OPTIMISATION OF TRAFFIC STEERING FOR HETEROGENEOUS MOBILE NETWORKS

by

SANDRA FREI

A thesis submitted to University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing, Electronics and Mathematics

In collaboration with Darmstadt Node of the CSCAN Network

March 2015

Optimisation of Traffic Steering for Heterogeneous Mobile Networks

Sandra Frei

Abstract

Mobile networks have changed from circuit switched to IP-based mobile wireless packet switched networks. This paradigm shift led to new possibilities and challenges. The development of new capabilities based on IP-based networks is ongoing and raises new problems that have to be tackled, for example, the heterogeneity of current radio access networks and the wide range of data rates, coupled with user requirements and behaviour. A typical example of this shift is the nature of traffic, which is currently mostly data-based; further, forecasts based on market and usage trends indicate a data traffic increase of nearly 11 times between 2013 and 2018. The majority of this data traffic is predicted to be multimedia traffic, such as video streaming and live video streaming combined with voice traffic, all prone to delay, jitter, and packet loss and demanding high data rates and a high Quality of Service (QoS) to enable the provision of valuable service to the end-user. While the demands on the network are increasing, the end-user devices become more mobile and end-user demand for the capability of being always on, anytime and anywhere. The combination of end-user devices mobility, the required services, and the significant traffic loads generated by all the end-users leads to a pressing demand for adequate measures to enable the fulfilment of these requirements.

The aim of this research is to propose an architecture which provides smart, intelligent and per end-user device individualised traffic steering for heterogeneous mobile networks to cope with the traffic volume and to fulfil the new requirements on QoS, mobility, and real-time capabilities. The proposed architecture provides traffic steering mechanisms based on individual context data per end-user device enabling the generation of individual commands and recommendations. In order to provide valuable services for the end-user, the commands and recommendations are distributed to the end-user devices in real-time. The proposed architecture does not require any proprietary protocols to facilitate its integration into the existing network infrastructure of a mobile network operator.

The proposed architecture has been evaluated through a number of use cases. A proof-of-concept of the proposed architecture, including its core functionality, was implemented using the ns-3 network simulator. The simulation results have shown that the proposed architecture achieves improvements for traffic steering including traffic offload and handover. Further use cases have demonstrated that it is possible to achieve benefits in multiple other areas, such as for example improving the energy efficiency, improving frequency interference management, and providing additional or more accurate data to 3rd party to improve their services.

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

2G	2nd Generation
3G	3rd Generation
3GPP	3rd Generation Partnership Project
4G	4th Generation
AAA	Authentication, Authorisation and Accounting
ABC	Always Best Connectivity
	Always Best Connected
ABS	Almost Blank Subframe
ACK	Acknowledgement
AES	Advanced Encryption Standard
AF	Application Function
AKA	Authentication and Key Agreement
AMBR	Aggregate Maximum Bit Rate
ANDSF	Access Network Discovery and Selection Function
ANPI	Average Noise Power Indicator
ANQP	Access Network Query Protocol
ANS	Access Network Selection
AP	Access Point
API	Application Programming Interface
APN	Access Point Name
APSD	Automatic Power Save Delivery
ARP	Allocation and Retention Priority
AUT	Automated

AVP	Attribute Value Pair
BBERF	Bearer Binding and Event Reporting Function
BCCH	Broadcast Control Channel
BER	Bit Error Rate
BPSK	Binary Phase-Shift Keying
BR	Black Rider
BS	Base Station
BSS	Basic Service Set
BSSID	Basic Service Set Identifier
Cap	Capability
CCA	Credit Control Answer
CCR	Credit Control Request
CDMA2000	Code Division Multiple Access 2000
CDN	Content Delivery Network
CM	Connection Manager
CoS	Class of Service
CRE	Cell Range Extension
CRUD	Create, Read, Update, Delete
CSG	Closed Subscriber Group
CTTC	Centre Tecnològic de Telecomunicacions de Catalunya
DB	Database
DGI	Data Gathering Interfaces
DiagMon	Diagnostics and Monitoring
DM	Device Management
DS	Distribution System
DSCP	Differentiated Services Code Point

DSMIPv6	Dual Stack Mobile IPv6
EAP	Extensible Authentication Protocol
EAP-AKA	EAP-Authentication and Key Agreement
EAP-SIM	EAP-Subscriber Identity Module
EAP-TLS	EAP-Transport Layer Security
EAP-TTLS	EAP-Tunnelled Transport Layer Security
EARFCN	Evolved Universal Terrestrial Radio Access Absolute Radio Frequency Channel Number
EBI	Evolved Packet System Bearer Identity
eICIC	enhanced ICIC
eNB	evolved Node B
EPC	Evolved Packet Core
ePDG	evolved Packet Data Gateway
EPS	Evolved Packet System
ESS	Extended Service Set
ETSI	European Telecommunications Standard Institute
E-UTRA	Evolved Universal Terrestrial Radio Access
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
FCS	Frame Check Sequence
FE	Front End
feICIC	further enhanced ICIC
GAS	Generic Advertisement Service
GBR	Guaranteed Bit Rate
GCM	Google Cloud Messaging
GCSE	Gateway Control Session Establishment
GCST	Gateway Control Session Termination

GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GRE	Generic Routing Encapsulation
GSM	Global Standard/System for Mobile Communications
GSoC	Google Summer of Code
GTP	GPRS Tunnelling Protocol
HBM	Host-Based Mobility
HEC	High Energy Consumption
HeNB	Home eNB
HESS	Homogeneous Extended Service Set
HESSID	Homogeneous Extended Service Set Identifier
HetNet	Heterogeneous Network
HLR	Home Location Register
HSPA	High Speed Packet Data Access
HSS	Home Subscriber Server
HTTP	Hypertext Transfer Protocol
IARP	Inter-APN Routing Policy
ICIC	Intercell Interference Coordination
IE	Information Element
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IFOM	Internet Protocol Flow Mobility
IKEv2	Internet Key Exchange version 2
IM	Internet Protocol Multimedia
IMS	Internet Protocol Multimedia Subsystem
IMT-Advanced	International Mobile Telecommunications-Advanced

iOS	i Operation System
IP	Internet Protocol
IP-CAN	IP-Connectivity Access Network
IPI	Idle Power Indication
IPsec	IP Security
IPv4	IP version 4
IPv6	IP version 6
ISMP	Inter-System Mobility Policy
ISP	Internet Service Provider
ISRP	Inter-System Routing Policy
ITU-R	International Telecommunication Union Radiocommunication Sector
KPI	Key Performance Indicator
LAC	Location Area Code
LCI	Location Configuration Information
LEC	Low Energy Consumption
L-GW	Local Gateway
LIPA	Local IP Access
LLCP	Logical Link Control Protocol
LMA	Local Mobility Anchor
L-PGW	Local Packet Data Network Gateway
LTE	Long Term Evolution
MAG	Mobility Access Gateway
MAPCON	Multi Access Packet Data Network Connectivity
MBR	Maximum Bit Rate
MBReq	Modify Bearer Request
MBResp	Modify Bearer Response

MCS	Modulation and Coding Scheme
MEC	Medium Energy Consumption
MICS	Media Independent Command Service
MIES	Media Independent Event Service
MIH	Media Independent Handover
MIHF	Media Independent Handover Function
MIHS	Media Independent Handover Services
MIIS	Media Independent Information Service
MIMO	Multiple Input Multiple Output
MME	Mobility Management Entity
MMS	Multimedia Messaging Service
MN	Mobile Node
MNO	Mobile Network Operator
MO	Management Object
MPDU	Media Access Control Protocol Data Unit
MSE	Mobility Speed Estimation
MU-MIMO	Multi User-MIMO
NAI	Network Access Identifier
NBM	Network-Based Mobility
NFC	Near Field Communication
NFV	Network Functions Virtualisation
NGSON	Next Generation Service Overlay Network
NIC	Network Interface Card
Ns-2	Network Simulator 2
Ns-3	Network Simulator 3
NSWO	Non-Seamless WLAN Offload

NW	Network
nwI_BR DB	network interface BR DB
OID	Object Identifier
OMA	Open Mobile Alliance
OMA DM	OMA-Device Management
OPIIS	Operator Policies for IP Interface Selection
OPNET	Optimized Network Engineering Tool
OS	Operation System
PBA	Proxy Binding Acknowledgement
PBU	Proxy Binding Update
pcap	Packet Capture
PCC	Policy and Charging Control
PCEF	Policy and Charging Enforcement Function
PCell	Primary Cell
PCID	Physical Cell ID
PCRF	Policy and Charging Rules Function
P-CSCF	Proxy-Call Session Control Function
PDB	Packet Delay Budget
PDN	Packet Data Network
PDN GW	PDN Gateway
PELR	Packet Error Loss Rate
PLMN	Public Land Mobile Network
PMIPv6	Proxy Mobile IPv6
PoA	Point of Attachment
PoD	Point of Decision
QAM	Quadrature Amplitude Modulation

QCI	QoS Class Identifier
QoS	Quality of Service
QUIC	Quick User Datagram Protocol Internet Connections
RACH	Random Access Channel
RADIUS	Remote Authentication Dial-In User Service
RAM	Random Access Memory
RAN	Radio Access Network
RAT	Radio Access Technology
RB	Radio Bearer
RCPI	Received Channel Power Indicator
RFC	Request For Comments
RNC	Radio Network Controller
RRC	Radio Resource Control
RRM	Radio Resource Management
RSCP	Received Signal Code Power
RSNI	Received Signal to Noise Indicator
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSS	Received Signal Strength
RSSI	Received Signal Strength Indication
RTS	Request to Send
S1-AP	S1-Application Protocol
SaMOG	Study on S2a Mobility based on GTP and WLAN access to the EPC network
SAP	Service Access Point
SCell	Secondary Cell

SCTP	Stream Control Transmission Protocol
SDF	Service Data Flow
SDN	Software Defined Networking
se	Signalling effort
SGSN	Serving GPRS Support Node
SGW	Serving Gateway
SIM	Subscriber Identity Module
SINR	Signal to Interference plus Noise Ratio
SIPTO	Selected IP Traffic Offload
SMS	Short Message Service
SNMP	Simple Network Management Protocol
SNR	Signal-to-Noise Ratio
SON	Self-Organising Network
SSDL	Signal Strength in the Downlink
SSID	Service Set Identifier
SSP	Subscription Service Provider
STA	Station
SU-MIMO	Single User-MIMO
TAC	Tracking Area Code
TDD	Time Division Duplex
TFT	Traffic Flow Template
ToS	Type of Service
TR	Technical Report
TS	Technical Specification
tse	Total signalling effort
TSF	Timing Synchronisation Function

TSMI	Traffic Steering Management Interface
UAM	Universal Access Method
UDC	User Data Convergence
UDP	User Datagram Protocol
UDR	User Data Repository
UE	User Equipment
UI	User Input
UMTS	Universal Mobile Telecommunications System
URL	Uniform Resource Locator
USIM	Universal Subscriber Identity Module
UTRAN	Universal Terrestrial Radio Access Network
VLR	Visitor Location Register
VNF	Virtual Network Function
VoIP	Voice over IP
WAN	Wireless Area Network
WFA	Wi-Fi Alliance
WiMAX	Worldwide Interoperability for Microwave Access
WISPr	Wireless Internet Service Provider roaming
WLAN	Wireless Local Area Network
WLANSP	WLAN Selection Policy
WPA2	Wi-Fi Protected Access 2
XML	Extensible Markup Language
XMPP	Extensible Messaging and Presence Protocol

Acknowledgements

I want to thank my supervisory team for their support.

Especially I want to thank my Director of Studies, Prof. Dr. W. Fuhrmann, for the support and the discussions which led to many new ideas. Your advice and guidance on both research as well as on my further way to go have been invaluable for me. Thank you for being my mentor and that I had the chance to learn so much from you.

I would also like to thank my supervisor Dr. B. V. Ghita for your constructive words, when I needed them and for your advice, guidance, and your valuable feedback on papers and the thesis. My thanks go also to my supervisor Prof. Dr. A. Rinkel. Without you, I would not have even started this long, exciting journey.

Many thanks to Carole Watson and Lucy Cheetham for your support by helping me with administrative issues.

I thank my family, especially my mother and my friends for the patience you have shown towards me, your understanding, when I once again had no time for you and your kindness and support throughout the years.

Without all of you, this would not have been possible. Thank you for your support.

Author´s Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Committee.

Relevant scientific seminars and conferences were regularly attended at which work was often presented, and several papers were prepared for publication.

Word count of main body of thesis: 53’779

Signed _____

Date _____

1 Introduction

The moving towards packet-oriented networks, where the Internet Protocol (IP) plays an important role, leads to new possibilities and also introduces new user requirements for the network. Specifically, the increasing use of multimedia services, like voice and video streaming or live streaming, creates new requirements for the network. Since these services are real-time services, consisting of time critical data, which are prone to delay, jitter and packet loss, end-to-end Quality of Service (QoS) has to be granted to enable valuable real-time services to the end-user. On the other hand the users are getting more mobile and want to be able to use services anywhere and anytime. The mobility mechanisms have to be able to provide quick handover so that the session still remains and the QoS does not degrade noticeably for the user, which is critical for real-time services. These expectations demand complex QoS and bearer management mechanisms and flexible mobility management from the network to provide a valuable service to the user. The Cisco forecast for global mobile data traffic predicts that global mobile data traffic will increase nearly 11 fold between 2013 and 2018 and that the number of mobile connected devices will exceed the world´s population by 2014 (Cisco, 2014). Globally, mobile data traffic has been approximately doubling each year for the last few years. The increasing number of smartphones, tablets but also sensors and machine-to machine modules are causing a massive mobile data traffic increase. The mobile communications industry is preparing to cope with an enormous increase of 1000x of mobile traffic by 2020 compared with 2010. Heterogeneous Radio Access Networks (RANs) with varying cell sizes are key elements to cope with this huge traffic volume. Beside the infrastructure, intelligent and smart traffic steering mechanisms are of paramount importance to distribute traffic among the heterogeneous RANs and offload heavy loaded RANs as well as parts of the core network.

The 3rd Generation Partnership Project (3GPP) has standardised the Evolved Packet System (EPS) which defines the Evolved Packet Core (EPC) network as well as the Long Term Evolution (LTE) Radio Access Technology (RAT) and the interworking of the EPS with the former defined technologies like Global System for Mobile

Communications (GSM), General Packet Radio Service (GPRS) and Universal Mobile Telecommunications System (UMTS) as well as the interworking with the 3GPP2 standards based on the Code Division Multiple Access 2000 (CDMA2000) technology, which is mostly used in the USA and Asian countries.

Since the Release 8 of the 3GPP standards, the integration of non-3GPP access technologies has been introduces which brings up new challenges for the handover, because inter-system handovers between 3GPP and non-3GPP access technologies are possible. In former 3GPP releases, inter-system handovers were limited to 3GPP defined technologies only.

The heterogeneous environment of the 3GPP and non-3GPP RANs is very complex. The main difficulty is the different realisation of the QoS, mobility, and bearer management within the various access technologies, as well as the different points of control: terminal-controlled and network-assisted, network-controlled and terminal-assisted, or a combination of these. To move towards a mobile network convergence it is necessary to optimise the traffic steering amongst heterogeneous networks including the inter-system handover and the offload mechanism. Traffic has to be distributed with intelligent traffic steering mechanisms among the available RANs to prevent access and core networks from traffic overload and to provide the best service towards the end-user.

1.1 Aims and Objectives of the Research

The objectives of this research is to provide access network independent optimisations for seamless mobility and QoS management with an intelligent and smart traffic steering framework for heterogeneous mobile networks to prevent networks from traffic over-utilisation and to provide valuable network services to the end-user.

An advanced architecture enabling novel traffic steering possibilities in heterogeneous mobile networks is proposed to tackle the problem of high traffic capacity while improving the QoS and supporting seamless mobility. Roaming cases are not considered in this thesis in order to reduce the complexity of this already complex research area.

The main objectives of the research can be summarised as follows:

1. Analysis of the current state of heterogeneous mobile networks focusing on traffic steering solutions as well as on mobility and QoS mechanisms.
2. Based on the analysis, define design decisions for the research.
3. Design the network architecture to provide new traffic steering possibilities with QoS and mobility optimisations. It should be feasible to integrate new evolving technologies into this architecture.
4. Identification of valuable use cases that the new architecture can be applied to.
5. Evaluation of the feasibility of the architecture by means of the proven use cases and by implementing parts of the architecture and functionalities in a simulation environment and evaluating the received results.

1.2 Thesis Structure

Chapter 2 contains definitions of the most important terms used throughout this thesis.

Chapter 3 consists of the related work covering the current defined 3GPP defined architecture for mobile networks with the discussion of actual mechanisms and functionalities of traffic steering as well as the main non-3GPP solutions for traffic steering. Finally the future trends of mobile networks are outlined briefly.

The design selections on options for protocols and network functions for the network architecture are made and defined in chapter 4.

The proposed architecture, called Black Rider (BR), is illustrated in chapter 5 by showing selected capabilities and by introducing the enablers for the BR concept.

The BR is described in detail in chapter 6 containing the BR building blocks and the functional cooperation between them, the data gathering mechanisms, the distribution of the commands and information and the location of the BR in the cloud.

In chapter 7 use cases are identified, discussed and partially evaluated, where the BR leads to improvements.

The main interesting functions of the BR were implemented in a simulator and the simulation results are presented in chapter 8. Therefore, the simulation setup is defined and the results are evaluated and discussed.

Finally, chapter 9 contains a summary of the achievements and the conclusions as well as the limitations and based on these, directions for further research are proposed.

2 Terminology

There are many specific technical terms defined and used in the context of the 3GPP EPS that are also used in this thesis. To prevent a misinterpretation of the terms used in this thesis, they are clearly defined in the following.

- **Massive increase of data traffic.** Recent years have witnessed an exponential growth of traffic in mobile environments. According to the Cisco global mobile data traffic forecast 2013 -2018 [1], mobile data traffic is expected to grow nearly 11-fold increase by 2018 over 2013. And the mobile communications industry is preparing to cope with a 1000x increase of traffic by 2020 over 2010. This tremendous increase of data traffic will affect the whole EPS, and this data volume can only be handled with solutions that relieve the core networks as well as the access networks from these tremendous traffic loads.
- **Traffic steering** is the process of distributing traffic among the available access networks and parts of the core network and even offload the core network completely from some types of traffic. Through traffic steering the traffic load can be balanced among cells and cell layers of the same or different RATs. Traffic steering can be performed statically or dynamically. Policies are generally used for traffic steering, for example, thresholds can be defined which, when a value falls below or exceeds a defined threshold, raises events that result in appropriate actions.
- **Traffic offloading** is the process of relieving the core network or RANs from traffic overload. Relieving the RANs from excess traffic can be achieved by moving the complete or parts of the traffic or traffic flows from one RAN to another available RAN. Depending on the selected solution, the offloading to another RAN can also relieve the core network. Since 3GPP Release 10 it is allowed to establish simultaneous connections over 3GPP and non-3GPP access.
- **A Handover** is the process when an end-user device changes the radio access node. Intelligent handover processing may lead to an offload from access or core networks through load balancing and traffic distribution. Handovers are

distinguished between inter-system handover and intra-system handover. An inter-system handover is processed when an end-user device changes from one RAT to a different RAT while the handover is called intra-system handover when the end-user device changes the radio access node while the RAT remains the same.

- **Always Best Connectivity / Connected (ABC).** The principle of Always Best Connectivity, also called Always Best Connected (ABC) is a vision that the end-user device is connected anytime and everywhere, with whatever mobility and speed to the most appropriate RAN in order to satisfy the requested service requirements. Different stakeholders have to be involved such as end-user, end-user device capabilities, network capabilities, requested services etc. End-users should be agnostic of the heterogeneity of the underlying RANs. This implies that the RAN selection is done without the assistance of the end-user.
- The terms **seamless** and **non-seamless mobility** apply to the handover as well as to the offloading process. Seamless mobility is evident for real-time or nearly real-time traffic which is prone to delay and jitter. Seamless mobility means, that the session is not interrupted and thus session continuity is provided. Non-seamless mobility may involve session interruptions during offloading or handover and therefore it is destined for nomadic and non-mobility usage. Nevertheless, non-seamless mobility can help to reduce the utilisation of access and core networks if traffic is not prone to jitter and delay. As long as the traffic from 3GPP and non-3GPP access goes through the EPC, the session continuity and therefore the seamlessness is ensured because the mobility anchor is the Packet Data Network Gateway (PDN GW) or the Gateway GPRS Support Node (GGSN) and an appropriate mobility protocol, which supports IP preservation, is applied.
- The terms **network-controlled**, **terminal-controlled**, **network-assisted**, and **terminal-assisted** are often used within the area of traffic steering indicating whether the network or the terminal or a mix of both owns the control. In all 3GPP defined RANs up to 3GPP Release 11 there is only one control mode: network-controlled and terminal-assisted. Within the 3GPP specified networks the network owns the full control of actions of the end-user devices. This implies also that the decision making processes remain in

the network and the end-user device will receive commands from the network. Terminal-assisted means that the terminal, the end-user device, performs either measurements in periodical intervals or on request of the network or configured through events that triggers the measurement procedure at the User Equipment (UE). But since non-3GPP access networks, such as Wireless Local Area Network (WLAN) is one, can be integrated into 3GPP´s networks, the situation has changed significantly and the Point of Control (PoD) is not as clear as in pure 3GPP network environments. A WLAN for example is not controlled by the network, instead it is terminal-controlled and all the decisions are made by the end-user device.

These terms are not sufficient to clearly differentiate the variants of control modes. Figure 1 tries to clarify the different variants of control modes. For every control mode the PoD is included to clearly identify the kind of control mode. Without the addition of the Point of Decision it would not be possible to differentiate the control mode variants a) and c) in Figure 1.

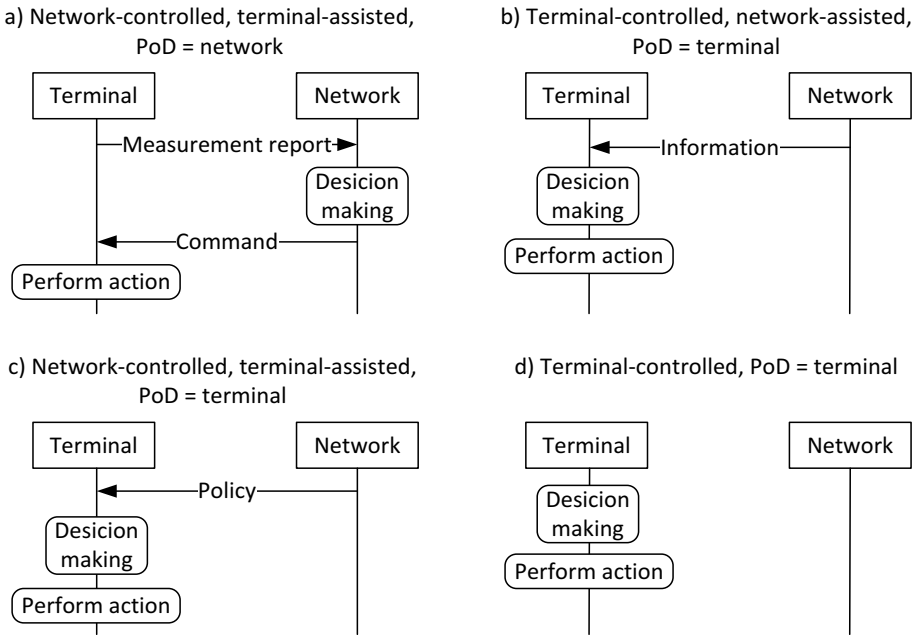


Figure 1: Control-modes including the Point of Decision

3 Related Work

All related papers considered in the following, that contain forecasts indicate that the increase of traffic cannot be handled by today´s mobile networks. Therefore, mechanisms have to be considered to cope with the predicted high increase of traffic volume to prevent access as well as core networks from overload. In the following, related work on standards as well as on scientific publications is presented to give a detailed overview of the existing solutions preventing networks from overload and from excessive high utilisation levels.

3.1 EPS Architecture

The 3GPP standardised the EPS which contains the definitions of the EPC network as well as the Evolved Universal Terrestrial Radio Access Network (E-UTRAN) and the interworking of the EPS with the former defined technologies like GSM, GPRS and UMTS as well as the interworking with the 3GPP2 standards where the CDMA2000 technology is defined, which is mostly used in the USA and Asian countries. Furthermore, the EPS standards consist of the definitions to integrate non-3GPP access technologies, such as WLAN and Worldwide Interoperability for Microwave Access (WiMAX) into the EPC.

Figure 2 shows the general view of the next generation mobile networks containing the three functional layers: application, services, which may contain the IP Multimedia Subsystem (IMS) defined in the Technical Specification (TS) (3GPP TS 23.228 V12.3.0, 2013), and transport, which consists of the EPS. The top layer is the application layer, which offers applications provided either by the operators or third parties. The service layer enables IP Multimedia (IM) services and defines the necessary functional network elements and mechanisms.

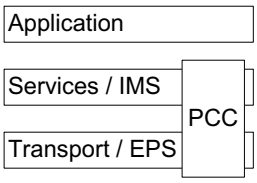


Figure 2: Overall view of the next generation mobile network

The EPC is the core network of the 4th Generation (4G) network. The EPS is based on IP transport and provides integration of multiple different RATs, such as the LTE, non-3GPP access technologies such as WiMAX or WLAN, as well as the integration of second and third generation 3GPP networks. Figure 3 shows a simplified architecture of the EPS. Not all interfaces and components, which belong to the EPS, are considered in this figure.

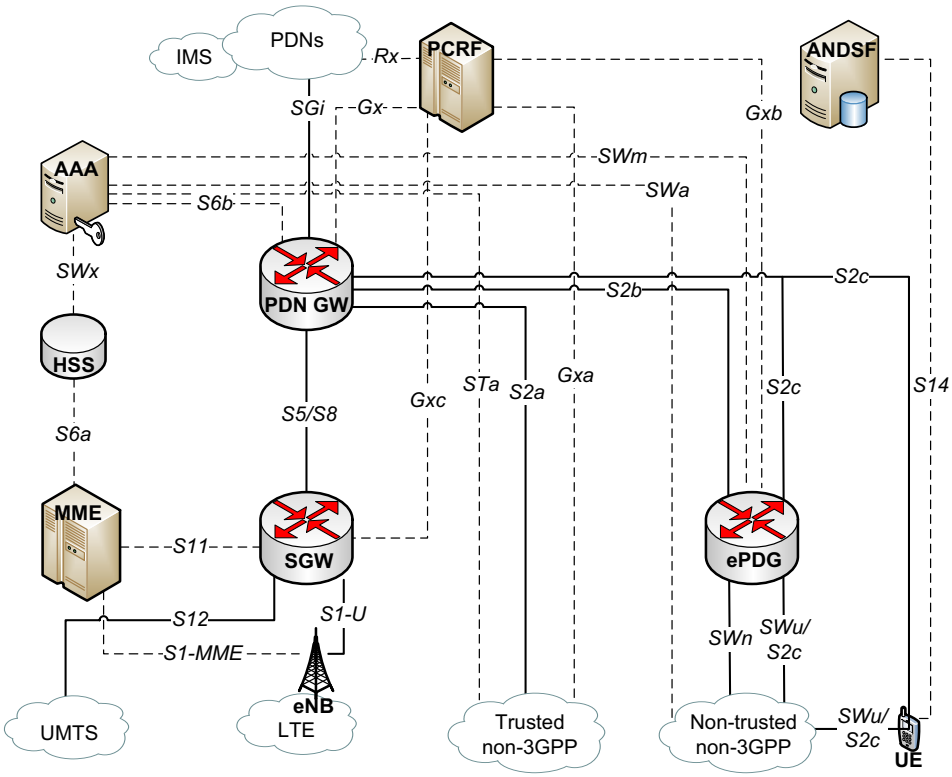


Figure 3: Simplified EPS architecture

The depicted components in Figure 3 are briefly introduced in the following. On the left side, there is the UMTS access network which is a 3rd Generation (3G) technology. The 3G core network components are not depicted in Figure 3. Instead the newly introduced core network components of the EPC are shown. The Authentication, Authorisation and Accounting (AAA) Server is used for security reason by all access networks. The Home Subscriber Server (HSS) is a database where all subscriber related information is stored. The Mobility Management Entity (MME) is a pure control plane element which is used to manage the bearers, as well as the mobility in 3GPP networks. Bearers are information transmission paths of a defined QoS between two network nodes. The bearer management consists of establishing, modifying, and deleting bearers. The two gateways in the middle, the PDN GW and

the Serving Gateway (SGW), are responsible for the routing of data traffic, to enforce the QoS concepts, and they represent mobility anchor points. These two gateways are often referred to as user plane elements because they carry data traffic from the user. Besides that, they are involved in the bearer management. The PDN GW provides access to other Packet Data Networks (PDNs) and it is the mobility anchor point for 3GPP and non-3GPP RATs. The evolved Node B (eNB) is located in the access network and provides the UE access to the core network. The policy rules related to QoS are provided through the Policy and Charging Rules Function (PCRF). The PCRF is also responsible for the authorisation of the QoS resources. The Access Network Discovery and Selection Function (ANDSF) provides environment information of neighbour cells and RATs to the UE to improve the handover and offload decisions. This service is available for both 3GPP and non-3GPP access networks. Beside environment information, the ANDSF provides different types of policies to the UE. The functionality of the ANDSF is discussed in more detail in 3.4.1.

The mobility protocol of the 2nd Generation (2G) and 3G 3GPP networks is the GPRS Tunnelling Protocol (GTP), which was adjusted and improved when used in the UMTS network and once more for the use within LTE/EPC. 3GPP defined for the EPC an alternative mobility protocol on the S5/S8 interface between the SGW and the PDN GW, which is the Proxy Mobile IP version 6 (PMIPv6) specified in the Request For Comments (RFC) (IETF RFC 5213, 2008). Another signalling protocol, mainly used for accessing databases and providing rules in the form of policies to the specific network elements, is an adapted version of the Internet Engineering Task Force (IETF) diameter base protocol (IETF RFC 3588, 2003), (3GPP TS 29.212 V12.3.0, 2013), (3GPP TS 29.214 V12.2.0, 2013).

The Release 8 of the 3GPP standards enables the integration of non-3GPP RATs into the EPC network and therefore all the services provided through the EPC are also available in the non-3GPP access networks. The non-3GPP RANs are divided into trusted (operator's own access network, or access network of partner operators) and non-trusted (external) networks. Non-trusted access networks are connected to the EPC through the evolved Packet Data Gateway (ePDG). The ePDG provides Extensible Authentication Protocol (EAP) method for 3G Authentication and Key Agreement (AKA) (EAP-AKA) (IETF RFC 5448 2009) authentication and data traffic protection through Internet Key Exchange version 2 (IKEv2) / Internet Protocol Security (IPsec).

An example technology of such a non-3GPP and non-trusted network is WLAN, but it is not limited to a specific access technology. These security related protections are not required for trusted networks and therefore the access towards the EPC is provided directly from a network component belonging to the access network which is comparable to the SGW from the EPC. The mobility protocols used with non-3GPP RANs are the GTP, thePMIPv6 and the Dual Stack Mobile IPv6 (DSMIPv6) protocol. The mobility protocols are described and analysed in more details in sections 3.2.3 and 4.3.

In order to provide IP multimedia services, the transport layer has to establish suitable bearers with appropriate QoS for the transport of multimedia data traffic. The bearer establishment is done through additional signalling between the service and transport layer by using the Policy and Charging Control (PCC) framework defined in (3GPP TS 23.203 V12.3.0, 2013) acting as a mediator between the service and the transport layer.

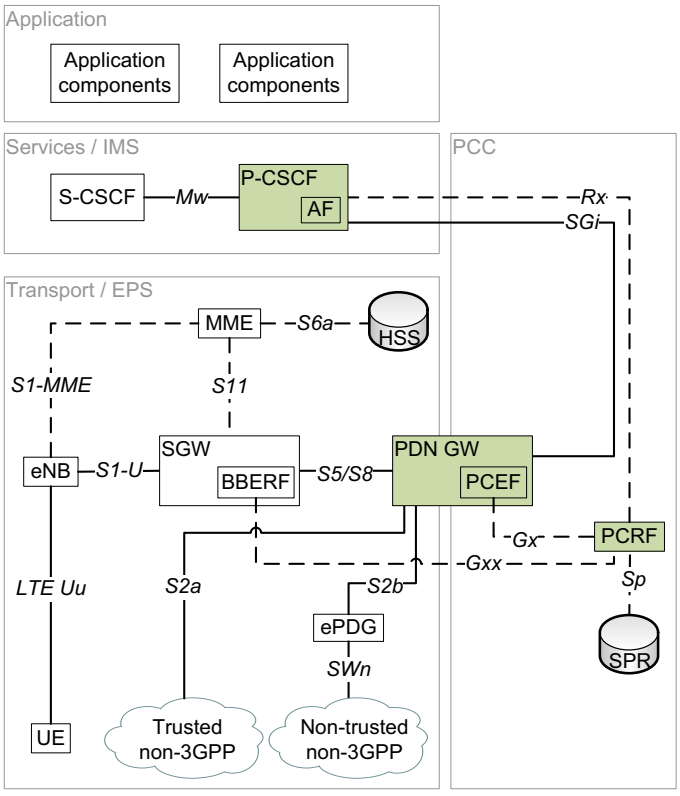


Figure 4: Detailed view of the next generation mobile network

Three main components are involved in communication between the service layer and the transport layer: the Application Function (AF), located at the Proxy-Call Session Control Function (P-CSCF) of the IMS, as well as the PCRF and the Policy and

Charging Enforcement Function (PCEF) which are located in the EPS network. This relationship is also shown in Figure 4, where a more detailed view of the next generation mobile network is given. The PCRF coordinates the establishment of IP-Connectivity Access Network (IP-CAN) bearers by processing incoming bearer establishment requests from the P-CSCF located in the IMS. An EPS bearer can carry multiple IP-CAN bearers. The PCC contains the two main functions:

- Charging the Service Data Flows (SDFs) through Online and Offline Charging System
- Policy Control which includes QoS control and signalling, gating control etc.

Since the focus of this research is mainly related to mobility and bearer management including QoS, the charging of the SDFs is out of scope and is therefore not considered in this thesis.

3.1.1 QoS Provisioning in the EPS

QoS is realised in the EPS through the so called EPS bearers. An EPS bearer consists of two identified end-points, an applied QoS class per EPS bearer and filters that describe the traffic flow to be able to filter the appropriate traffic for the EPS bearer. The EPS bearer is established in the EPC with GTP tunnels and on the air interface with the radio bearer. EPS bearers can only be established if GTP is the selected mobility protocol, since the alternative mobility protocol, the PMIP, is not able to carry filters which are mandatory to establish an EPS bearer.

3.1.2 Bearer Level QoS Parameters

The EPS bearer is a basic element of the QoS definitions in the EPS network. One EPS bearer does only have one level of QoS, but can comprise one or multiple IP-CAN bearers. As a result of this, all IP-CAN bearers which are transported over the same EPS bearer have to support the same QoS. An EPS bearer is associated with a QoS Class Identifier (QCI), an Allocation and Retention Priority (ARP) and an authorised bit rate for both up- and downlink direction (3GPP TS 23.203 V12.3.0, 2013). The QCI is a scalar to classify traffic in a technology independent way. There are 9 QCI values, each of which has a defined priority, an upper bound for the average Packet Delay Budget (PDB) between the UE and the PCEF, and a maximum Packet Error Loss Rate

(PELR) defined between the UE and the radio base station. Four QCI values have a Guaranteed Bit Rate (GBR), the other five are used for non-GBR bearers. If the EPS bearer is a GBR it is associated with a GBR and a Maximum Bit Rate (MBR) value. The ARP value defines the priority of bearers including the following:

- a value from 1 to 15, where 1 is the highest priority
- the pre-emption capability flag to indicate, if the SDF can get resources that are already assigned to another SDF
- the pre-emption vulnerability flag to indicate, if the assigned resources of the SDF can be allocated to another SDF with a higher priority.

The ARP values can be used for admission control as well as for rejection decision in case of congestion or handover situations. To enforce traffic shaping for non-GBR EPS bearers, the Aggregate Maximum Bit Rate (AMBR) value is used, which defines a maximum bit rate for a group of non-GBR bearers. The AMBR value can be set per Access Point Name (APN), see equation (1), or per UE, see equation (2), and is enforced by both, the PDN GW and the UE, to protect the bit rate of the GBR bearers.

$$\text{APN-AMBR} = \sum_{i=1}^{n} BR\ of\ non-GBR\ bearer_i\ \ associated\ with\ APN_x \qquad (1)$$

$$\text{UE-AMBR} = \sum_{i=1}^{n} BR\ of\ non-GBR\ bearer_i\ \ associated\ with\ UE_x \qquad (2)$$

To ensure that the service level is enforced the AMBR and MBR values can be used for traffic shaping in the UE, the eNB and the PDN GW. The resources of an EPS GBR bearer are allocated through the bearer establishment procedure or are rejected because of a lack of resources or if the QoS cannot be accepted for example due to the type of subscription. But if a GBR EPS bearer once is established the resources are allocated and guaranteed.

3.1.3 Policy and Charging Control

The PCC mediates the cooperation between the service and transport layer. It is responsible for deciding on the allocation and authorisation of the required bearer resources on request of the service layer and provides session admission control in

cooperation with the transport layer. The focus of this research is QoS and mobility management and therefore the charging system of the PCC is not considered.

The policy control provides the following functions:

- Binding, i.e. the generation of an association between an SDF and the appertaining IP-CAN bearer.
- Admission control of SDFs.
- Event reporting, for example, notification for triggering bearer modifications.
- QoS control, i.e. the authorisation and enforcement of the maximum QoS that is authorised for an SDF or an IP-CAN bearer.
- IP-CAN bearer establishment

The binding mechanism of the PCC associates the application with the transport layer session providing the required QoS parameters within the applied rules. The binding mechanism consists of three steps:

1. **Session binding**. The session binding associates the service layer session of the AF with the corresponding transport layer session, the IP-CAN session. This assignment is performed by the PCRF using, e.g., the following IP-CAN parameters: UE IP address, UE identity, information about the accessed PDN.
2. **PCC and QoS rule authorisation**. The PCC and QoS rule authorisation is used to determine the QoS level by selecting QoS parameters for PCC and QoS rules. The authorisation of the PCC and QoS rules results in the decision, if the user is granted access to the required service or not, and defines the constraints. The decision is performed by the PCRF.
3. **Bearer binding**. The bearer binding comprises of the association between PCC and QoS rules and an IP-CAN bearer within an IP-CAN session. Bearer binding involves the decision, whether an existing IP-CAN bearer can be used or a new IP-CAN bearer has to be established.

The PCC rules applied for the binding mechanism can either be dynamic or predefined. The dynamic PCC rules are generated in the PCRF and are provisioned towards the PCEF where they are installed and finally enforced. The predefined PCC rules are preconfigured in the PCEF. The PCRF is able to activate or deactivate a preconfigured PCC rule or even a set of PCC rules.

3.2 Traffic Offload

Traffic offload is a method of coping with the expected huge increase of traffic and can be used to offload specific RANs as well as parts or the whole of the core network from specific traffic. In the following different types of offloading are considered.

3.2.1 Non-Seamless WLAN Offload

The most elementary way of offloading traffic supported by the EPC is the Non-Seamless WLAN Offload (NSWO). These days, it is usual for UEs to be equipped with 3GPP 2G/3G/4G interfaces as well as with an IEEE 802.11 WLAN interface. The UE connects to WLAN access when it is available. Principally, the WLAN and 2G/3G/4G connections can be operated in parallel, but today, the common way is that the UE disconnects the 2G/3G/4G data connection when the UE has a connection via the WLAN access. This may result in bad user experience if the WLAN access is overloaded or provides only a poor data rate. The UE acquires a local IP address on the WLAN interface. The traffic routed via the WLAN interface is not routed through the EPC, but bypasses it. If the UE moves out of the WLAN coverage the IP connectivity is lost and therefore there is no service continuity.

3.2.2 Local IP Access and Selected IP Traffic Offload

With 3GPP Release 10 Local IP Access (LIPA) and Selected IP Traffic Offload (SIPTO) were introduced in (3GPP TR 23.829 V1.3.0, 2010). Both offloading mechanisms are used to offload traffic from the core network due to the use of a local breakout point. Therefore, these two offloading mechanisms do not offload traffic from RANs. The basic features of both, LIPA and SIPTO, have been standardised in 3GPP Release 10-12. There is an ongoing 3GPP study of mobility with both offloading mechanisms.

LIPA is used to offload traffic through 3GPP Home eNBs (HeNB) (femtocells) to local IP networks such as residential and enterprise networks or intranets. Every device, located in the local network, as well as every network accessible by the local network can be accessed through LIPA. Figure 5 shows the LIPA architecture for the case where the Local Gateway (L-GW) is collocated with the HeNB.

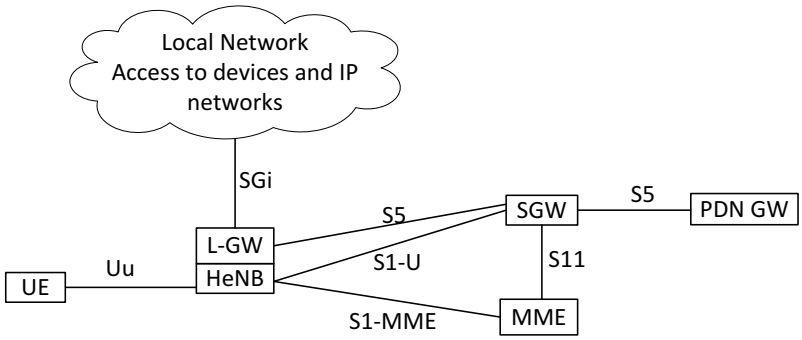


Figure 5: LIPA architecture where the L-GW is collocated with the HeNB (Rel. 10)

The signalling is still done through the core network, but the user plane traffic destined for the local network does not cross the core network anymore. A simultaneous connection through LIPA and the core network is possible. The L-GW supports partial PDN GW functions such as IP address assignment for UEs, policy-based packet filtering and rate policing and shaping per UE. Furthermore, SGW downlink data buffering functionality is supported by the L-GW. But the L-GW supports neither any paging functionality nor mobility management functions. This results in an inability of the L-GW to page the UE. Therefore, the S5 interface is needed between the L-GW and the SGW. If data traffic is received at the L-GW it is buffered and a dummy packet is sent to the SGW to trigger the paging. The SGW sends a data notification message to the MME which triggers the MME to start the paging procedure. As soon as the connection to the L-GW is established, the L-GW starts sending the buffered data towards the UE. In 3GPP Release 10 the L-GW is collocated with the HeNB and mobility of the LIPA PDN connection is not supported. The connection is lost if the UE moves from one HeNB to another.

In 3GPP Release 11 the standalone L-GW is introduced. There are two variants of LIPA architectures with a standalone L-GW: LIPA with a standalone L-GW with control via the SGW and the MME and LIPA with a standalone L-GW where the Sxx interface is deployed between the L-GW and the HeNB.

Figure 6 shows the LIPA architecture where the interface between the L-GW and the HeNB only supports the user plane traffic and the control plane signalling is done through the EPS core network entities, the SGW and the MME, by using the same interfaces, the S1-MME (with the S1-Application Protocol (S1-AP) as the overlaid protocol), the S11 (with GTP-C as the overlaid protocol), as defined in the EPS with the use of eNB, and the S5 interface between the L-GW and the SGW to trigger the

paging procedure. The paging is done through the path including the network elements: L-GW – SGW – MME – HeNB. LIPA session continuing during mobility between Local Network and macro network can be achieved through the core network signalling.

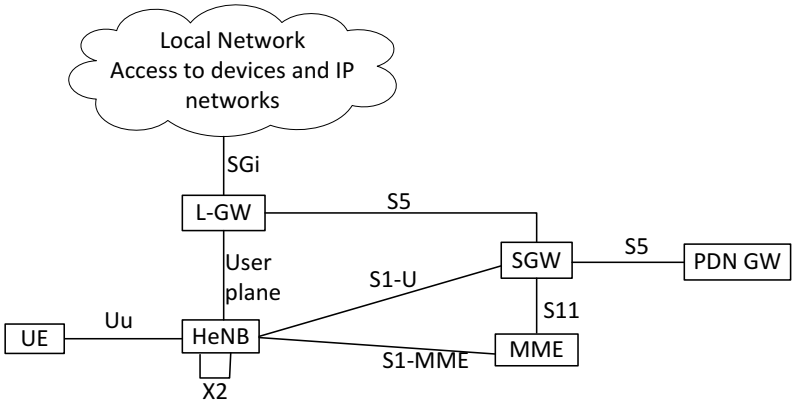


Figure 6: LIPA architecture with standalone L-GW with control via the SGW and the MME

Figure 7 shows the LIPA architecture with a standalone L-GW where the Sxx interface is deployed between the L-GW and the HeNB. The Sxx interface supports both user plane and control plane traffic. The details of the Sxx are not yet defined in the Technical Report (3GPP TR 23.859 V12.0.1, 2013). But it is assumed in the Technical Report (TR) that GTP will be chosen as the mobility protocol on the Sxx interface. The interfaces S1-U is still necessary to send and receive data plane traffic to the EPC. The S11 and S1-MME interfaces transport the control plane traffic. The S5 interface between the L-GW and the SGW is still used to trigger paging to the SGW because the L-GW does not support paging mechanisms.

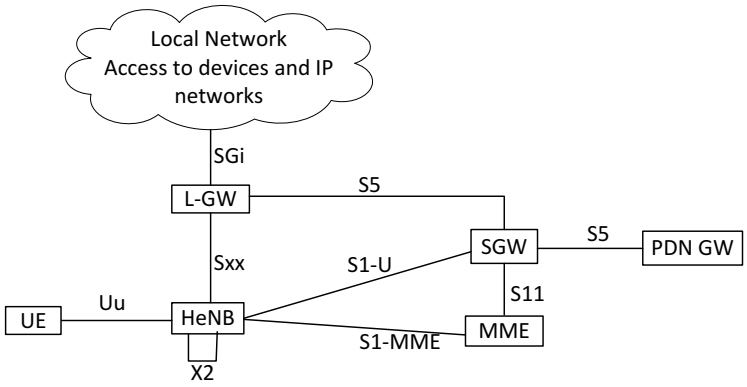


Figure 7: LIPA architecture with standalone L-GW and Sxx interface between L-GW and HeNB

SIPTO is used to offload internet traffic through selective routing. SIPTO is applied to eNBs and HeNBs. Selected IP traffic is therefore routed through either the most

optimal path in an operator´s core network or bypassing the core network completely. The simultaneous traffic offloading through SIPTO and traffic routed through the EPC is allowed. While the breakout point for LIPA is always at the residential/enterprise/local network, 3GPP Release 10 defined the breakout point of SIPTO at or above the RAN. The architecture is shown in Figure 8. Instead of routing the traffic to the PDN GW, traffic is routed to the geographically nearer and therefore local PDN Gateway (L-PGW). With the SIPTO above RAN architecture macro mobility within macro networks and between macro networks and (H)eNBs is supported.

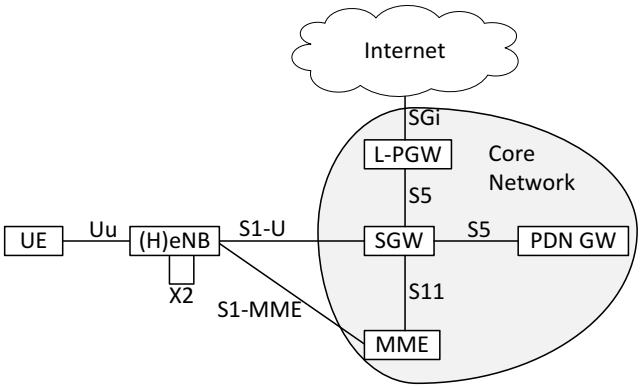


Figure 8: SIPTO above RAN

In 3GPP Release 12, an additional breakout point with SIPTO is defined in the residential/enterprise/local network. It is called SIPTO at the local network and it is different from LIPA. Beside user consent and user awareness, it is indicated in the technical report (3GPP TR 23.859 V12.0.1, 2013) that it is expected that SIPTO at the local network needs to be supported for small cells and not only for (H)eNB systems. Several architectural solutions exist for SIPTO at the local network. Two of them are selected to be adopted in the technical report (3GPP TR 23.859 V12.0.1, 2013). These are SIPTO above RAN architecture in local network and SIPTO at the local network reusing Release 10 LIPA architecture.

The architecture shown in Figure 9 reuses the SIPTO above RAN architecture for the use within local networks. This solution is an operator-controlled service and as a result of this, the RAN and gateway functionality is within operator-controlled environment and considered as trusted. This solution is for the non-collocated case where the (H)eNB is not collocated with another network entity. With the SIPTO above RAN architecture in local network macro mobility as well as mobility between local network and macro network is supported.

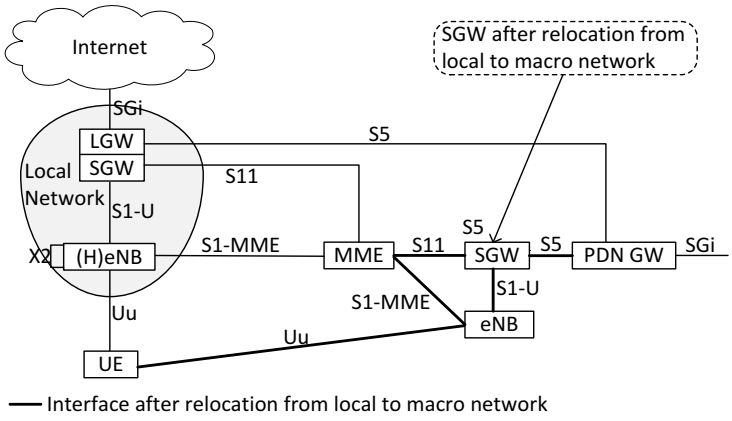


Figure 9: SIPTO above RAN architecture in a local network

The architecture shown in Figure 10 reuses the 3GPP Release 10 LIPA architecture for SIPTO at the local network. The L-GW is collocated with the HeNB. The signalling is performed through the core network components, the SGW and the MME. For 3GPP Release 12, SIPTO at the local network does not support PDN connection session continuity of IP data sessions, when the UE moves away from the (H)eNB. The SIPTO at the local network PDN connection can be re-established via the PDN re-establishment procedure when the UE moves away from HeNB.

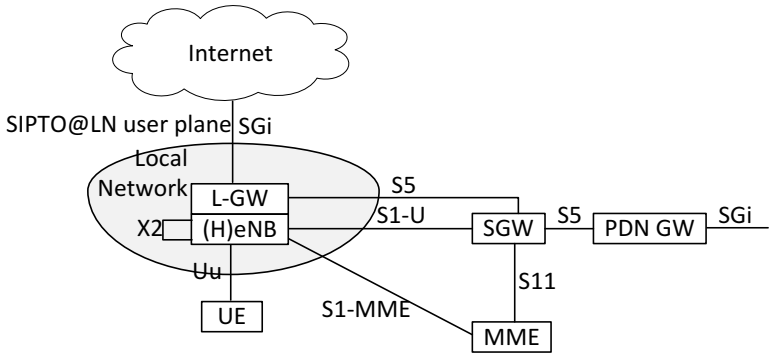


Figure 10: SIPTO at the local network reusing Release 10 LIPA architecture

The Technical Report (3GPP TR 23.859 V12.0.1, 2013) analyses the mobility for LIPA and SIPTO at the local network. In (3GPP TR 23.859 V12.0.1, 2013) it is proposed, that mobility is supported for both LIPA architectures as well as for the SIPTO at the local network architecture with the standalone L-GW acting as the anchor point of the LIPA/SIPTO connection. Thus, the UE can maintain an IP connection while changing the HeNB (LIPA/SIPTO) and the eNB (SIPTO), if the target (H)eNB has IP connectivity towards the L-GW, which is the anchor point of the LIPA/SIPTO connection. Figure 11 illustrates this mobility situation. Because the L-GW acts as an anchor point for mobility, continuous LIPA/SIPTO IP connectivity may principally be supported within

a local (H)eNB network, where all the (H)eNBs have an IP connectivity towards the same L-GW. But 3GPP Release 12 does not include mobility yet, as this study is still a Technical Report and not yet a 3GPP standard. According to (3GPP TR 23.859 V12.0.1, 2013) SIPTO at the local network is expected to be supported also for eNBs small cells and not only for HeNBs femtocells.

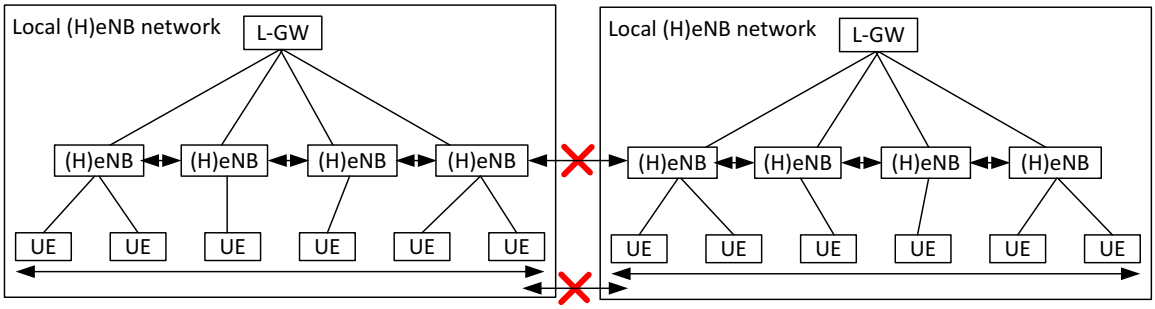


Figure 11: LIPA/SIPTO mobility with standalone L-GW

In the technical report (3GPP TR 23.859 V12.0.1, 2013) it is concluded that the solution architecture with the standalone L-GW for LIPA mobility will be adopted as the baseline architecture. These architectures are shown in Figure 6 and Figure 7 and can either contain the control plane via the SGW and the MME or the Sxx interface supporting both, user plane and control plane. Furthermore, in the TR (3GPP TR 23.859 V12.0.1, 2013) are conclusions on SIPTO that the following two architectures will be adopted as architectures: For L-GW and SGW collocated it will be the architecture shown in Figure 9. For a standalone L-GW it will be the architecture depicted in Figure 10.

3.2.3 Multi Access PDN Connectivity and IP Flow Mobility

Multi Access PDN Connectivity (MAPCON) and IP Flow Mobility (IFOM) are investigated in (3GPP TR 23.861 V1.7.0, 2012). Since it is a technical report, the specification process is not yet finalised. Since 3GPP Release 10 an end-user device is allowed to maintain 3GPP and non-3GPP accesses simultaneously. MAPCON and IFOM are quite similar to each other. Both offload mechanisms have in common that they require the UE to have multiple interfaces and the ability to run them simultaneously to enable simultaneous access to the EPC through 3GPP and non-3GPP RANs. The kind of interfaces and access technologies has a huge impact on how non-3GPP RANs are integrated into the EPC. On one hand there is the trusted non-3GPP access and on the other hand there is the non-trusted non-3GPP access. The

kind of access type (trusted or non-trusted) makes a big difference as to how they are integrated into the EPC, as can be seen in Figure 12. Trusted non-3GPP RANs have direct access via the S2a interface to the PDN GW, whereas non-trusted non-3GPP RANs have to establish a security tunnel to the ePDG in order to get access to the PDN GW via the S2b interface.

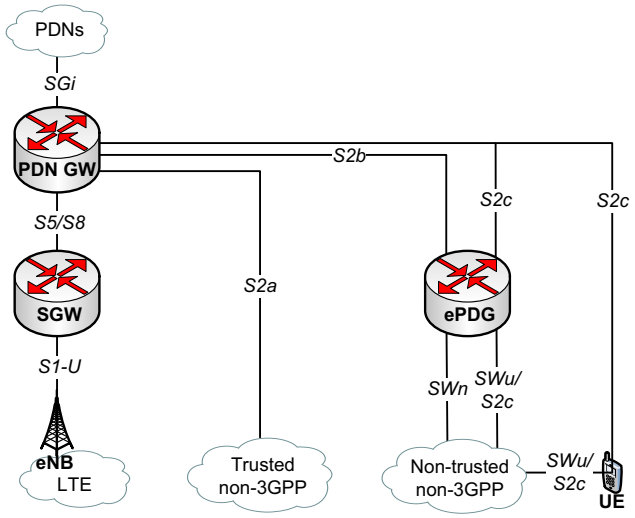


Figure 12: EPS architecture for non-3GPP access technologies

The three interfaces to connect a non-3GPP access technology towards the EPC are S2a, S2b, and S2c interfaces. Table 1 indicates which interfaces are used to provide access to the PDN GW for trusted and non-trusted access technologies.

Table 1: Type of non-3GPP access support by the S2x interfaces

	S2a	S2b	S2c
Trusted	X		X
Non-trusted		X	X

Table 2 indicates, which protocols are defined on which interfaces and of what kind (i.e. Host-Based Mobility (HBM) and Network-Based Mobility (NBM)) the protocol is.

Table 2: Mobility protocol support of the S2x interfaces

	S2a	S2b	S2c	HBM	NBM
PMIPv6	X	X			X
GTP	X	X			X
DSMIPv6			X	X	

PMIPv6 and GTP are both network-based mobility protocols whereas the DSMIPv6 is a host-based mobility protocol. PMIPv6 was the first mobility protocol that was allowed to be used on the S2a interface to connect a UE over a trusted non-3GPP

access technology towards the PDN GW (EPC) and on the S2b interface to connect non-trusted non-3GPP access technologies towards the PDN GW (EPC). In the Technical Report (3GPP TR 23.834 V10.0.0, 2010), GTP is also allowed as an alternative mobility protocol on the S2b interface to connect a UE over a non-trusted non-3GPP access network towards the EPC. The use of GTP on the S2a interface connecting trusted non-3GPP access technologies towards the PDN GW (EPC) was defined in (3GPP TR 23.852 V12.0.0, 2013). As a result of the development in the field of WLAN security in the past years, WLAN is considered as trusted non-3GPP access. GTP tunnels can be used for connection, if GTP is selected on the S2a or S2b interface. In case PMIPv6 is used on the S2a or S2b interface Generic Routing Encapsulation (GRE) tunnels are used. It does not matter if PMIP or GTP is applied to the S2b, between the UE and the ePDG the connection is secured with an IPsec tunnel established over the SWu interface. If DSMIPv6 is used on the S2c interface with a trusted access network there is a direct DSMIPv6 tunnel between the UE and the PDN GW. If DSMIPv6 is used on the S2c interface with a non-trusted access network there is an IPsec tunnel between the UE and the ePDG and a DSMIPv6 tunnel between the UE and the PDN GW. The IPsec tunnel between the UE and the ePDG ensures that each end-user device can communicate with the network in a secure way. This creates a logical association between each end-user device and the ePDG over the SWu interface, which carries signalling and user data. The SWn interface between the non-trusted access network and the ePDG transports all the signalling and data traffic from the SWu interface.

MAPCON makes it possible to route different active PDN connections through different access networks. The MAPCON enabled UE has usually a 3GPP and one trusted or non-trusted non-3GPP interface. MAPCON can be realised using the S2x interfaces. As mentioned before, it is necessary that both interfaces can be run simultaneously. Figure 13 shows the situation where the UE is equipped with an LTE (3GPP access) and a WLAN (non-3GPP, trusted access) interface. The UE has a connection to the PDN 1, the IMS, and to the PDN 2, the internet. Each PDN connection consists of one IP flow. The Voice over IP (VoIP) flow is routed via the LTE access to the IMS, whereas the web flow is routed via the WLAN to the internet. Every flow is routed through the PDN GW. If a new IP flow, destined to the IMS, is added, this newly added IP flow has to be routed via the LTE access, because it is prohibited

to route connections destined for the same PDN over different accesses. PDN connections, containing all associated IP flows, can be moved seamlessly from 3GPP access to non-3GPP access and vice versa. Therefore, it is possible, for example, in the case where the UE moves out of the coverage area of the WLAN access network, that all the flows belonging to the PDN 2 (Internet) are seamlessly moved to the 3GPP access.

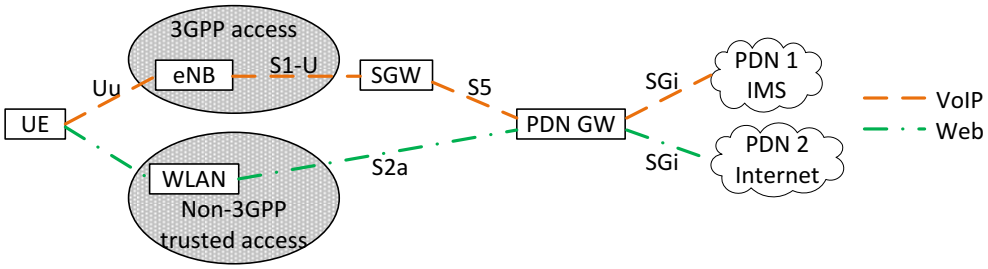


Figure 13: Multi Access PDN Connectivity (MAPCON) with trusted WLAN

IFOM allows to route different IP flows to the same PDN connection through different access networks. This means, that in contrast to MAPCON, IFOM provides an additional level of granularity for inter-system mobility, because IFOM does not have this limitation that all flows belonging to the same PDN have to be routed through the same access network. Therefore, it is possible to route flows, destined for the same PDN, through different access networks. This situation is shown in Figure 14 with a VoIP flow routed via the LTE access network and a non-conversational video stream routed via the WLAN access network. Both IP flows are destined for the IMS PDN. Like MAPCON, IFOM is realised using the S2x interfaces as well. Individual IP flows can be moved from 3GPP access networks to non-3GPP access networks and vice versa seamlessly and without limitations.

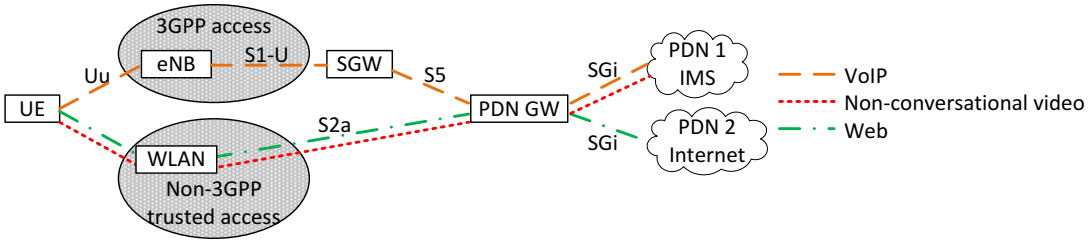


Figure 14: IP Flow Mobility (IFOM) with trusted WLAN

As mentioned before, the Technical Report (3GPP TR 23.861 V1.7.0, 2012) is not finalised yet. But the following service requirements are defined in (3GPP TR 23.861

V1.7.0, 2012) and these requirements give an idea of what can be expected from MAPCON and especially from IFOM:

- *Service continuity should be provided when the UE moves from the 3GPP access to non-3GPP access and vice versa.*
- *If the UE is under the coverage of more than one access, including 3GPP and non-3GPP accesses, it should be possible for the UE to communicate using multiple accesses simultaneously, if the UE is authorised by subscription to access all of the involved PDNs and all of the involved access networks.*
- *It should be possible to select one access when a flow is started and re-distribute the flows to/from a UE between accesses while connected.*
- *It should be possible for the operator to enable and control the simultaneous usage of multiple accesses.*
- *It should be possible to distribute flows to/from a UE between available accesses based on the characteristics of the flows and the capabilities of the available accesses, subjected to user's preferences and operator's policies. For example, when both 3GPP and non-3GPP accesses are available, flows with high QoS requirements (e.g. voice) may not be routed through the non-3GPP access, in order to prevent loss of service.*
- *It should be possible for the operator to define policies for the control of the distribution of IP flows between available accesses. Each policy shall include a list of preferred accesses and whether the policy may be overridden by the user's preferences.*

Furthermore, it is defined in (3GPP TR 23.861 V1.7.0, 2012) that the study of the GTP-based S2a support for trusted WLAN access with seamless offload and flow mobility is deferred until the study on S2a mobility based on GTP and WLAN access to the EPC network (SaMOG) Release 12 study (3GPP TR 23.852 V12.0.0, 2013) is completed.

3.3 Small Cells and Heterogeneous Networks

The predicted huge increase of traffic makes the development of small cells necessary and the term small cells is recently used widely in combination with handover and offloading. The Small Cell Forum, formerly known as the Femto Forum, defined small cells as follows: *'Small cells' is an umbrella term for low-powered radio access nodes that operate in licensed and unlicensed spectrum and typically have a range of 10 metres to several hundred metres. These contrast with a typical mobile macro cell that might have a range of up to several tens of kilometres. The term covers femtocells, picocells, microcells, metrocells and public Wi-Fi.*

Small cells are operator controlled. In Figure 15 an overview about the usage of the different cell types for different operational areas is provided. Femtocells have the smallest range followed by picocells that have a medium range and metrocells and microcells that have the largest range.

Nevertheless, the pico-, micro-, and metrocells can be based on femtocell technology. In general pico-, micro-, and metrocells are deployed by operators, whereas femtocells are usually user deployed. Femtocells differ also from the other types of cells because the access to a femtocell is limited to members of a so called Closed Subscriber Group (CSG) while the access to the other cell types is open and only restricted through the subscriber access permission of the operator.

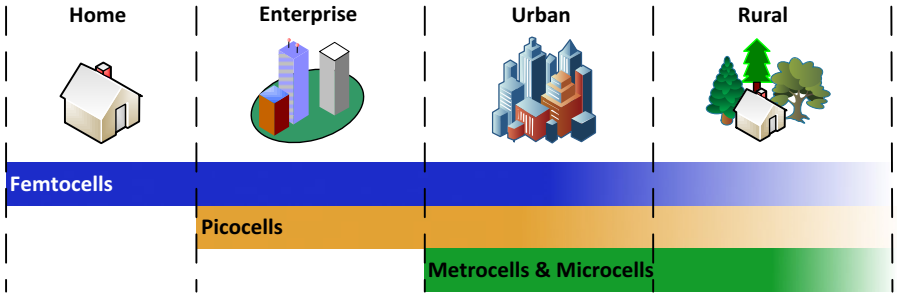


Figure 15: Operational area of different types of cells

These small cells are considered promising, able to cope with the predicted mobile traffic volume increase by offloading traffic from macro cells to small cells. In general the principle is very simple. The bigger the cell is, the less traffic capacity can be provided per unit area because all the users in the cell share the available radio resources. If small cells are used the cell coverage is smaller and therefore the cell capacity is shared among fewer users. As a result, the traffic capacity per unit area is higher with small cells than with macro cells. The drawback of small cells is that there are more handovers when the end-user device is moving around and seamless mobility is provided. In the 3GPP Technical Report (3GPP TR 36.839 V11.1.0, 2012) amongst other things, the number of handover failures in small cells has been analysed by evaluating simulation results of several companies that used uniquely defined simulation parameters. The results showed that the speed of the UE has a significant impact on the handover performance. The trend of these small cell simulation results indicates that UEs with a high speed suffer from much higher handover failure rates than UEs with a low speed. This is true for all analysed simulation scenarios in (3GPP TR 36.839 V11.1.0, 2012). The considered speeds were

3 Km/h, 30 Km/h, 60 Km/h, and 120 Km/h. For UEs at low speed (i.e. speed < 30 Km/h), no significant problems have been observed in terms of handover failure and loss of connectivity. Furthermore, the findings of the (3GPP TR 36.839 V11.1.0, 2012) were that in general the handover performance is better the lower the load is. Therefore, it has been concluded in (3GPP TR 36.839 V11.1.0, 2012) that the addition of picocells while the overall system load remains constant may have a positive effect on the mobility performance, because handover failures/radio link failures are reduced due to a reduction of the load per cell. Furthermore, this study shows that if macro and small cells are deployed together as Heterogeneous Networks (HetNets) the mobility performance is not as good as in macro-only networks. Handovers from pico to macro cells perform worst. The mobility performance in a HetNet environment gets worse if the speed of the UE is increased. In the first LTE Release (Release 8) the Mobility Speed Estimation (MSE) was introduced to classify the UEs by different mobility states. The UEs count the number of past cell crossings which results in a roughly estimation of their handover rate and then map it to a defined mobility state, which indicate the speed of the UE: low, medium or high. Handover parameters are then adjusted according to the mobility state. However, with a HetNet deployment there is a mixture of small and macro cells and the mobility estimation based on the number of dense small cell crossings is not equal to the number of macro cell crossings. Therefore, it was concluded in (3GPP TR 36.839 V11.1.0, 2012) that the currently applied mobility state estimation by the UE is not accurate enough and thus has to be enhanced.

In general, small cells are deployed with indoor and outdoor hotspots and at the cell edge of macro cells to improve the performance at the macro cell edge, as can be seen in Figure 16.

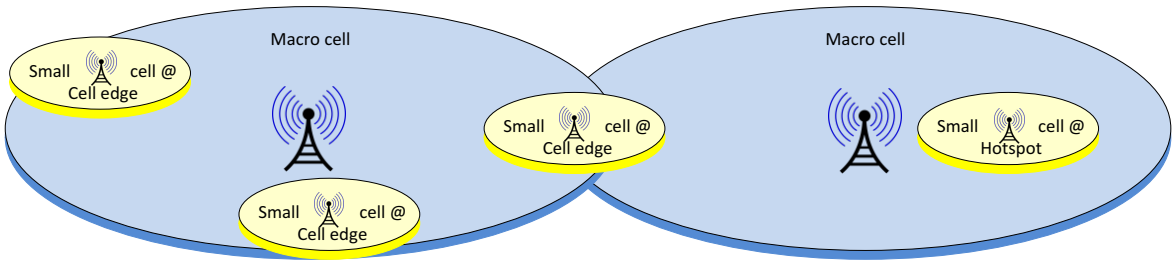


Figure 16: Typical HetNet deployment

One of the features of small cells is that they are low power nodes, that means that the transmit (Tx) power of a small cell is lower than the Tx power of a macro cell, such as the eNB is one. From the introduction of small cells the HetNets have emerged in 3GPP Release 10 by combining macro- and small cells. HetNets consists of macro cells overlaid with small cells. The RATs of macro and small cell can be either different or similar as well as the used carrier frequencies.

3.3.1 Interference Protection Between LTE Cells

The LTE technology is designed to apply a frequency reuse of one. This causes interference at the cell edge. To cope with the intercell interferences of neighbour cells Intercell Interference Coordination (ICIC) was introduced in 3GPP release 8. ICIC is an optional method to decrease the interference between neighbour cells. Figure 17 shows a possible situation when ICIC is applied. The centre of the cells all use the same spectrum but with low power, so that the neighbour cells are not affected by interferences of the $f_{spectrum}$. The frequencies f1 to f3 are placed in a frequency reuse pattern at the edge of the cells and therefore higher power is used.

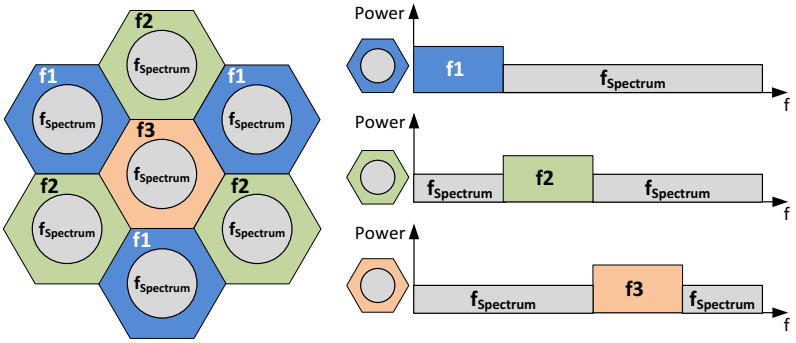


Figure 17: ICIC mechanism

The necessary information between neighbour cells is exchanged via the X2 interface.

3.3.2 Interference Protection for LTE-Advanced HetNets

ICIC is only suitable for preventing interferences with neighbour LTE cells, but interference prevention with HetNets is not covered. Therefore, in 3GPP Release 10 LTE-Advanced specification (3GPP TS 36. Series, 2011) the enhanced ICIC (eICIC) has been introduced. The eICIC defines an additional time dimension to ensure orthogonality in the time domain for UEs within overlapping cells. Therefore, the Almost Blank Subframes (ABSs) are introduced in the eICIC. ABSs are selected

subframes that are not used by the aggressor cell to send user data in the downlink direction, instead only control channels and cell-specific reference signals are transmitted with reduced power. The intercell interference caused by the aggressor cell to the victim cell is reduced in ABS and therefore these ABSs can be used by victim cells to send data. Figure 18 shows an ABS configuration at the aggressor cell. During the ABS no user data is sent from the aggressor cell instead the victim cell uses these ABSs to send data. Victim cells are allowed to send also data in non-ABSs. The interference within non-ABSs is higher, compared to the interference of ABSs, but the closer the UE is located to the small cell antenna the stronger the signal is and therefore it is possible to send data also in non-ABSs. Aggressor cells are either macro cells or small cells. A small cell located at the edge of a macro cell is an aggressor for UEs at the cell edge connected to the macro cell, because the small cell is interfering with the macro cell. A macro cell is an aggressor for UEs that are connected to small cells located near the centre of the macro cell, because the macro cell interferes with the small cell. The ABSs have to be configured in the aggressor cells and the frame structures are communicated over the X2 interface.

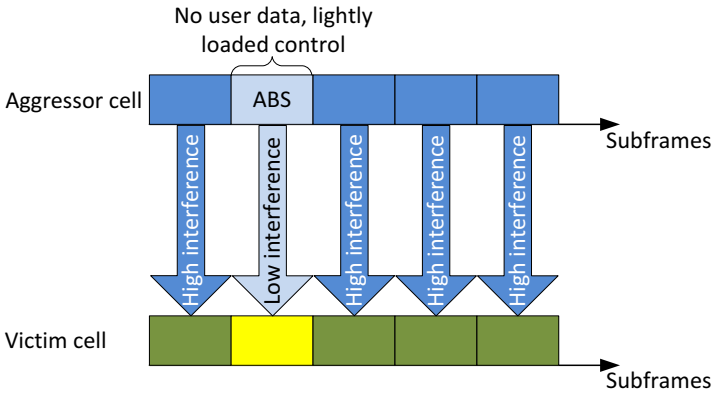


Figure 18: eICIC time domain mechanism

Another problem is the cell selection within HetNets. In general the UE measures the Signal Strength in the Downlink (SSDL) of different cells and based on these measurements the cell with the strongest signal strength is selected. In HetNet topologies this mechanism results in selecting the macro cell in most cases which lead to the problem of over-utilisation of macro cells and under-utilisation of small cells. As a result, the whole advantage of small cells cannot be used, because small cells transmit at very low power and this results in stronger signal strength for macro cells unless the UE is located very close to a small cell. To tackle this problem Cell Range

Extension (CRE) is used which adds an offset to the SSDL of a small cell. As a result small cells can gather UEs also at the cell edge depending on the used offset which is added to the actual SSDL value. The drawback of this approach is the increased interference in the downlink at the UEs in this CRE zone, because the signal strength of the macro cell is high and the signal strength of the small cell is rather low because the added offset to the SSDL. Therefore, the UEs in the CRE zone are predestined to make use of the ABSs in order to be able to send in subframes which have only little interferences.

The 3GPP Release 11 introduced a further enhanced ICIC (feICIC) mechanism including, for example, interference reduction at the UE and at the eNB. There are several specifications 3GPP is working on within 3GPP Release 12. These improvements are not considered further in this thesis since they do not affect the architecture of the proposed approach in this thesis. The most important fact for the proposed approach in this thesis is that LTE-Advanced HetNets can be deployed with a frequency reuse of one.

3.3.3 Wi-Fi Small Cells

The HetNets improvements discussed so far are destined for LTE-Advanced macro and small cells. But HetNets can also consist of LTE-Advanced macro cells and Wi-Fi used as the small cell technology. Qualcomm provided a comparison of HetNets with LTE as the RAT for macro and small cells and LTE as the RAT for macro cells and Wi-Fi as the RAT for small cells in (Qualcomm, 2011). For the Wi-Fi small cells RAT the Institute of Electrical and Electronics Engineers (IEEE) 802.11n standard (IEEE 802.11n, 2009) has been used. The HetNet with LTE-Advanced macro and picocells operates both at the same frequency; therefore, the frequency reuse is one. The eICIC mechanisms have been applied with the use of ABS and CRE and advanced receiver supporting interference cancellation at the UEs. There is no interference between LTE-Advanced cells and Wi-Fi cells because they operate in different frequency bands. Wi-Fi can interfere only with other Wi-Fi cells, but this is explicitly not considered in the analysis (Qualcomm, 2011). The comparison shows that LTE-Advanced macro and picocells consistently outperform LTE-Advanced macro and Wi-Fi small cells. The performance is measured with the throughput increase of HetNets in relation to the throughput of macro-only coverage. The performance difference is more pronounced

in suburban scenarios with relative low density of small cells with 200% better performance compared to 20% better performance in hotspot distribution scenarios. Reasons for these differences are the higher throughput of the LTE-Advanced picocells compared with Wi-Fi small cells and the eICIC mechanisms (ABS, CRE, and interference cancellation). Especially the cell range extension of LTE-Advanced picocells has the effect of more UEs being associated with LTE-Advanced picocells than with small Wi-Fi cells.

But since 2008, a new IEEE 802.11ac standard amendment (IEEE 802.11ac, 2013) has been worked out and ratified in December 2013. Even before the amendment was ratified, there were already 802.11ac devices on the market known as the first wave of 802.11ac products. Now the second wave of 802.11ac products are expected to hit the market. The 802.11ac amendment (IEEE 802.11ac, 2013) provides several improvements compared with (IEEE 802.11n, 2009). The main improvements are summarised in Table 3.

Table 3: Main 802.11ac improvements compared to 802.11n

	802.11n	**802.11ac**
Bandwidth	20, 40 MHz	20, 40, 80 160 MHz
Multiple Input Multiple Output (MIMO)	MIMO, Single User-MIMO (SU-MIMO)	MIMO, SU-MIMO and Multi User-MIMO (MU-MIMO)
Spatial streams	1-4	1-8
Beamforming	Optional	Optional
Modulation	16 and 64 Quadrature Amplitude Modulation (QAM)	16, 64, 256 QAM
Frequency band	2.4 and 5 GHz	5GHz
Max. Throughput	600Mbps	6.93 Gbps

Beamforming is the ability to focus radio frequency energy in a distinct direction to improve the delivery of individual stations. The 802.11n standard specified beamforming but not the way beamforming has to be implemented. This is due to the abundance of protocol combinations for beamforming available in the 802.11n standard. As a result, beamforming was implemented differently by vendors and the devices from different vendors are rarely compatible. Therefore, the market acceptance has been very poor. The 802.11ac standard restricts these plenty of possibilities for beamforming and enables a consistent implementation of the beamforming function. Therefore, it could be expected that beamforming might be

more widely adapted by vendors in future. Beamforming is the basis to enable Single User- Multiple Input Multiple Output (SU-MIMO) and Multi User-MIMO (MU-MIMO). Figure 19 shows the principle of SU-MIMO, where data is sequentially sent to the end-user devices.

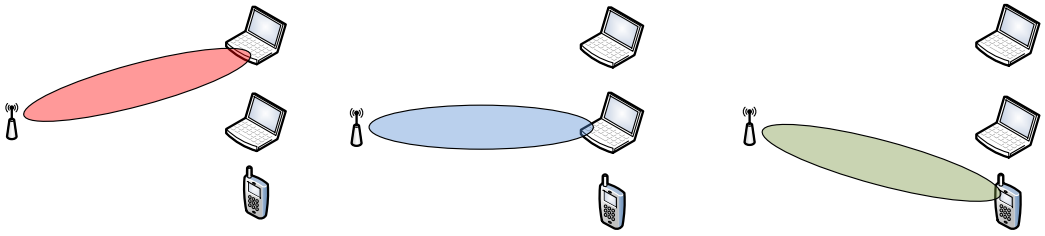


Figure 19: SU-MIMO Beamforming

Figure 20 shows the principle of MU-MIMO. Data can be sent to the end-user devices in parallel. Therefore, multiple end-user devices can be served in parallel by one access point.

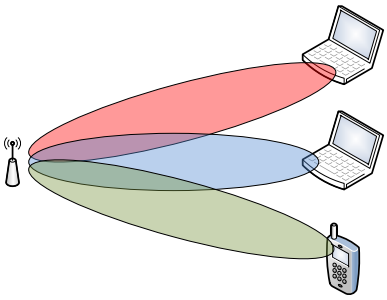


Figure 20: MU-MIMO Beamforming

The maximum range of the 802.11ac standard remains unchanged compared to the 802.11n standard in the 5GHz frequency spectrum. But the data rates increase with the 802.11ac standard as shown in Table 4 for the case of Single User MIMO. The blue rows are expected to be used with smartphones having up to two antennas. Broadcom presented a Wi-Fi 2x2 MIMO chip at the Mobile World Congress 2014 in Barcelona. It has been the first time a smartphone has more than one antenna and as a result, MIMO is possible. The orange rows are expected to be used with tablets and notebooks, because of the amount of 3 and 4 antennas. But the use of 4 antennas is nowadays not common in notebooks and tablets. The grey rows make use of the 160 MHz channel and 4 or more antennas. The use of 4 and more antennas is nowadays left to special setups and devices and not yet commonly available in notebooks and tablets. Also the use of a 160 MHz channel is not common, because only few 160 MHz

channels can be used in parallel, which is a problem with a high density of Wi-Fi access points. The minimum and maximum data rates in Table 4 are calculated with the Modulation and Coding Scheme (MCS) 0 and 7 (0= Binary Phase-Shift Keying (BPSK), 7=64 Quadrature Amplitude Modulation (QAM)) for the 802.11n standard and 0 and 9 (0= BPSK, 9=256 QAM) for the 802.11ac standard with the long guard interval for the minimum data rate and the short guard interval for the maximum data rate. The end-user devices are called Stations (STAs) in IEEE 802.11 WLAN standards.

Table 4: Selected data rates of the 802.11n and 802.11ac standards

[Mbps]				[MHz]
Min. data rate	**Max. data rate**	**Standard**	**SU-MIMO**	**Bandwidth**
13.00	144.00	802.11n	2x2	20
27.30	300.00	802.11n	2x2	40
27.30	400.00	802.11ac	2x2	40
29.25	433.00	802.11ac	1x1	80
40.95	450.00	802.11n	3x3	40
40.95	600.00	802.11ac	3x3	40
54.60	600.00	802.11n	4x4	40
58.50	867.00	802.11ac	2x2	80
58.50	867.00	802.11ac	1x1	160
117.00	1733.00	802.11ac	2x2	160
87.75	1300.00	802.11ac	3x3	80
117.00	1733.00	802.11ac	4x4	80
234.00	3470.00	802.11ac	4x4	160
468.00	6930.00	802.11ac	8x8	160

Smart phones
Tablets/Notebooks
Special use cases

Table 5 shows data rates for selected setups with MU-MIMO, which is only available in the 802.11ac standard. Since the end-user devices have a maximum of 2 antennas, these examples are all applicable for smart phones. MU-MIMO is only supported in the downlink direction according to (IEEE 802.11ac, 2013).

Table 5: Selected data rates with Multi User-MIMO

[Mbps]					[MHz]
Min. data rate per STA	Max. data rate per STA	Aggregated data rate	MU-MIMO	Number of STAs	Bandwidth
29.25	433.00	1732	4x1	4	80 MHz
58.50	867.00	3470	4x1	4	160 MHz
58.50	867.00	3470	8x2	4	80 MHz
58.50	867.00	3470	8x2	4	80 MHz

Smart phones, Tablets/Notebooks

The maximum reach of Wi-Fi 802.11ac signal remains unchanged compared to the 802.11n in the 5GHz frequency spectrum. Technology advances of the 802.11ac standard extend the data rate at every distance compared to the previous wireless standards 802.11n standard (XIRRUS, 2013) (Cisco, 2014). The reason for this behaviour of the 802.11ac standard is the robustness. While the channel width of the 802.11n standard is a maximum of 40MHz, the 802.11ac can use 80MHz or even 160MHz, and therefore it is possible to use the long guard interval and a lower MCS and the throughput is still equal or higher but more robust than the throughput and robustness of the 802.11n standard with the highest MCS and short guard interval applied.

With these improvements of the 802.11ac standards, the performance differences between LTE-Advanced and Wi-Fi small cells are expected to narrow. The analysed performance advantage of 200 % in suburban areas and 20% in urban areas with LTE-Advanced small cells over Wi-Fi small cells (Qualcomm, 2011) is no longer valid with the use of the 802.11ac standard. Besides that, Wi-Fi small cells deployment is cheaper than LTE-Advanced small cells. But the advantage that CRE brings to the LTE-Advanced small cells that more end-user devices can be served by LTE-Advanced small cells than by Wi-Fi small cells, still remains because there is no such mechanism defined for Wi-Fi.

3.4 Traffic Steering

Because of the tremendous increase of data traffic is predicted in the next years (Cisco, 2014) traffic steering is of eminent importance in order to cope with this problem. Whereas former 3GPP generations do not allow that non-3GPP access

networks get access to the 3GPP core network the 3GPPs 4G networks are able to integrate all kinds of access networks whether they are trusted or non-trusted, 3GPP or non-3GPP networks. Furthermore, the introduction of small cells led to the development of the HetNets with multiple cell layers, different cell sizes, multiple RATs, using different or identical carrier frequencies. Such kinds of networks are not completely new, i.e. in former 3GPP generations (<4G) there are also multiple cell layers, different cell sizes, multiple RATs (limited to 3GPP only access networks) using different carrier frequencies, but the integration of non-3GPP access networks into the 3GPPs core network was not supported and the 4G networks allow an increasing node density of low power nodes, that was not the case, to this extend, in <4G networks. All the involved RATs in <4G were 3GPP access technologies and therefore the RANs were all under control of the mobile network operators. All developments in the field of small cells and HetNets result in the demand for more sophisticated traffic steering mechanisms compared with < 4G networks. In the following subchapters standards and publications are references that survey existing and proposed solutions for traffic steering in HetNets.

To be able to distribute traffic appropriately, policies are used to enable network operators to manage traffic demands and provide at the same time an adequate QoS experience for the end-user. Thereby, the following, not exhaustive list, gives some major requirements that have to be considered:

- The dynamic load situation has to be taken into account to be able to distribute traffic in an efficient and reasonable way.
- The subscription level has to be taken into account. This implies different QoS levels per end-user.
- The enforcement of different QoS levels has to be supported through prioritisation of the traffic.
- The blocking of certain traffic has to be supported.
- The ability to offload traffic to other access networks is required.

3.4.1 Access Network Discovery and Selection Function

Before the 3GPP Release 8, static pre-provisioned policies have been used to manage and steer the handover and offload behaviour of the UEs. This is how most implementations today decide which access networks shall be used by the UE. The

3GPP Release 8 introduced the optional ANDSF to allow a more dynamic control of these policies. In the subsequent 3GPP Releases the ANDSF has been enhanced (3GPP TS 24.302 V12.3.0, 2013), (3GPP TS 23.402 V12.3.0, 2013). The ANDSF is designed to support all 3GPP and non-3GPP access networks. The ANDSF is intended to support the UE to perform network discovery and selection according to operator policies where access network solutions are not sufficient for this process. The information exchange between ANDSF and UE is based on the Open Mobile Alliance (OMA) Device Management (DM) (OMA, 2013a) and uses the ANDSF Management Object (MO) specified in the 3GPP standard (3GPP TS 24.312 V12.3.0, 2013) to manage the inter-system mobility policy and access network discovery information provided by the ANDSF. The information gained in form of MOs is stored in the UE. The ANDSF provides the following kinds of policies to UEs:

- Inter-System Mobility Policy (ISMP): This policy contains a set of operator-defined rules and preferences that have an impact on the handover decision of a UE that is not capable of connecting to the EPC through multiple access networks simultaneously. ISMPs could for example contain preferences, which RAN is optimal for EPC communication, as well as restrictions for inter-system handovers in order to allow or deny certain inter-system mobility. The inter-system mobility policy can be used for example to define the preferences of the operator to route all traffic over WLAN or cellular access.
- Access network discovery information: The ANDSF provides a list of access networks available for the UE, including the access type technology, RAN identifier (such as a Service Set Identifier (SSID) in the case of Wi-Fi), and other technology specific information like carrier frequencies in use.
- Inter-System Routing Policy (ISRP): The ANDSF provides inter-system routing policies to UEs. UEs that are capable of routing IP traffic simultaneously over multiple radio access interfaces process these policies. UEs that do not support this capability ignore these policies. The inter-system routing policies are used to route IP traffic specifically, such as restricting certain IP traffic flows and/or specific APNs from specific access technology types, or to realise the appropriate selection, by the UE by applying IP filter rules, of the access technology, access network, and APN. The inter-system routing policies consist of one or more IP filter rules that define the access technologies/access

networks which shall or shall not be used by the UE to route IP traffic that matches these IP filters. These kinds of policies are used to provide the operator´s preferences to UEs configured for MAPCON, IFOM or non-seamless WLAN offload.

- Inter-APN Routing Policy (IARP): IARP was first introduced in the Technical Report (3GPP, TR 23.853 V12.0.0, 2012). These policies can be either defined statically or dynamically. Only inter-APN capable UEs process these policies. UEs that are capable of routing IP flows simultaneously over multiple radio access interfaces where each interface is associated with a different APN are inter-APN capable UEs. These interfaces can be linked to different or the same access network. An inter-APN capable UE can use these policies to select an outgoing interface based on the preferred APN in the policies. Furthermore, these policies can include information for identifying IP flows that are applicable for non-seamless WLAN offload.
- WLAN Selection Policy (WLANSP): are defined in the standards (3GPP TS 23.402 V12.3.0, 2013) (3GPP TS 24.312 V12.3.0, 2013) and developed further in (3GPP TR 23.865 V12.1.0, 2013). The ANDSF provides the WLANSP information for the UE to select and reselect a WLAN access network. The policy can be divided into different WLAN criterion groups. If all criterions of a group are fulfilled by the WLAN this WLAN is classified as a preferred access network. If there is no preferred WLAN access, because the criterions are not fulfilled, it depends on the implementation how the UE selects a WLAN access network. The possible criterions are the following: PreferredRoamingPartnerList, MinBackhaulThreshold, MaximumBSSLoadValue, RequiredProtoPortTuple, PreferredSSIDList, and SPExclusionList. These criterions have been introduced in (3GPP TS 24.312 V12.3.0, 2013) and (3GPP TS 23.402 V12.3.0, 2013) and are defined in the Hotspot 2.0 specification (Wi-Fi Alliance Hotspot 2.0, 2012).

All types of policies provided by the ANDSF can have validity conditions, which indicate when provided policies or information are valid or not. These validity conditions can be associated for example with locations or time.

The policies are sent in a downlink from the ANDSF to the UE. But location information can be sent also on the uplink from the UE to the ANDSF, for example, the

UE includes its location information in the request to the ANDSF (3GPP TS 24.302 V12.3.0, 2013). The UE location provides the following types of location information (3GPP TS 24.312 V12.3.0, 2013):

- Geographical coordinates containing latitude and longitude values.
- Cellular area such as Public Land Mobile Network (PLMN), Tracking Area Code (TAC), Location Area Code (LAC), Cell identifier etc.
- WLAN location such as Homogeneous Extended Service Set Identifier (HESSID), SSID, Basic Service Set Identifier (BSSID).

Figure 21 shows the S14 interface between the UE and the ANDSF. The information exchange over the S14 interface is based on the OMA DM (OMA, 2013a) protocol. The OMA DM protocol is used to provide network policies in the downlink and information about the UE in the uplink. The information is contained in MOs. So called DM trees, collecting multiple MOs, are stored in the ANDSF and in the UE. To synchronise MOs with the OMA DM version 1.2 and 1.3 the Synchronization Markup Language (SyncML), an XML-based open standard protocol for synchronising mobile data independent of underlying networks, platforms and devices, is used. 3GPP defined their own MO called ANDSF MO (3GPP TS 24.312 V12.3.0, 2013) to manage the inter-system mobility policy and access network discovery information provided by the ANDSF. There exist the OMA DM versions 1.2, 1.3 and 2.0. The ANDSF MOs are compatible with the OMA DM protocol specifications of version 1.2 and upwards.

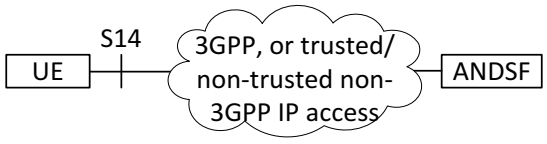


Figure 21: Interface to ANDSF

The latest OMA DM version 2.0 is not backwards compatible on the transaction level with former versions of the protocol, but the MOs are backward compatible. The OMA DM version 2.0 uses JavaScript Object Notation (JSON) for serialisation, a language independent and lightweight data format. OMA DM version 2.0 uses RESTful protocol principles, where the Representational State Transfer (REST) denotes an architectural style for networked applications (Fielding, 2000). REST is not limited to use the Hypertext Transfer Protocol (HTTP), but since the Web´s primary transfer protocol is HTTP, REST commonly uses HTTP as the transfer protocol. In combination

with HTTP, REST represents a lightweight alternative to the XML-based Simple Object Access Protocol (SOAP) and other remote procedure call techniques.

OMA DM includes diagnostics and monitoring specifications known under the term (DiagMon). This so called DiagMon enabler addresses the following areas according to the DiagMon V1.0 (OMA, 2011a):

- ***Diagnostics Policies Management:*** *Support for specification and enforcement of policies related to the management of diagnostics features and data.*
- ***Fault Reporting:*** *Enable the device to report faults to the network as the trouble is detected at the device.*
- ***Performance Monitoring:*** *Enable the device to measure, collect and report Key Performance Indicators (KPIs) data as seen by the device such as on a periodic basis.*
- ***Device Interrogation:*** *Enables the network to query the device for additional diagnostics data in response to a fault*
- ***Remote Diagnostics Procedure Invocation:*** *Enables management authorities to invoke specific diagnostics procedures embedded in the device to perform routine maintenance and diagnostics.*
- ***Remote Device Repairing:*** *Enables management authorities to invoke specific repairing procedures based on the results of diagnosis procedures.*

The most important DiagMon function in respect to traffic steering is the performance monitoring which provides the end-user device with the capability to measure, collect and report KPIs. These KPIs are of paramount importance when making appropriate traffic steering decisions.

3.4.2 Hotspot 2.0

The Hotspot 2.0 standard is defined by the Wi-Fi Alliance (WFA) and is based on the IEEE standards 802.11u (IEEE 802.11u, 2011) and the 802.11i, which is integrated within the 802.11-2007 standard (IEEE 802.11-2007, 2007). Hotspot 2.0 is the technology specification behind the Wi-Fi Alliance's Passpoint certification program. Hotspot 2.0 improves the ability of end-user devices to use nearby Wi-Fi access points, as well as to discover, associate, and authenticate to access points in an automatic way and without an intervention from the end-user. Hotspot 2.0 Release 1 provides the ability to the end-user device to use query mechanisms (IEEE 802.11u, 2011) to discover information about the available roaming partners and the type of credentials used with the access point. The query mechanism is very similar to the access network discovery mechanism known from the ANDSF. The protocol is defined in (IEEE 802.11u, 2011) and is called Access Network Query Protocol (ANQP)

and contain information elements such as 3GPP cellular network information, Network Access Identifier (NAI) realm list, roaming consortium list, domain name list, venue name, venue information, operator-friendly name, IP address-type availability, Wireless Area Network (WAN) metric (e.g. uplink/downlink estimation of speeds, loading, link status, and whether the WLAN is at capacity), connection capability, operating class, network authentication type, HESSID, access network type, internet available, and Basic Service Set (BSS) load. A Homogeneous Extended Service Set (HESS) is a collection of Basic Service Sets (BSSs) within the same Extended Service Set (ESS), where all external networks or Subscription Service Provider (SSP) are reachable by all the BSSs. The HESS is identified by the HESSID and consists of a BSSID of an access point belonging to the HESS. Which BSSID is configured as the HESSID is the responsibility of the administrator. It is important, that all BSSs belonging to the same HESS have the same HESSID configured. If the HESSID and the SSID are used in combination the specific service provider network can be identified, since the HESSID is a unique ID. In the standard (IEEE 802.11u, 2011) the SSP is introduced. The SSP is an operator which manages the user´s subscription and the associated credentials and offers connection to network services. An access point can provide access to multiple SSPs because of the various roaming agreements. As a result, the end-user devices can dynamically query which SSPs are available. Furthermore, the access credentials can be independent from the SSIDs and this enables the end-user devices to automatically discover roaming agreements on access points it has never been previously connected to. To enable the end-user device to query an access point´s capabilities and supported SSPs while unauthenticated and not associated with that access point the Generic Advertisement Service (GAS) is defined in (IEEE 802.11u, 2011). Public action frames are used to transport GAS related information and it is based on the request/response principle, where the end-user devices send the requests as defined by the Hotspot 2.0 Release 1 (Wi-Fi Alliance Hotspot 2.0, 2012). Beacon and probe request/reply frames are modified and extended in (Wi-Fi Alliance Hotspot 2.0, 2012) based on the definitions in (IEEE 802.11u, 2011) with different or additional information elements such as the interworking element, advertisement protocol, and roaming consortium. These additional information elements provide information about the following:

- The interworking element provides information about the type of network such as private or public, free or fee, access to the internet or not, venue info (Business, education, vehicular, industrial etc.), venue types (movie theatre, stadium, library, restaurant etc.), HESSID.
- The advertisement protocol information element is used to prevent the access point from broadcasting all the information to all STAs in the range of the access point. Therefore, the advertisement protocol information contains the protocol, which is supported to provide the advertisement information, such as ANQP, which is mandatory, or Media Independent Handover (MIH) etc.
- The roaming consortium is a group of service providers which have roaming agreements with one another to enable the end-user device to authenticate with the user´s credentials. The roaming consortium information element provides the end-user device with the information which roaming consortiums or service providers are available through the access point.

In the 802.11u standard (IEEE 802.11u, 2011) two additional information elements are introduced: the QoS map set information element to map layer 3 QoS priority of a service provider to layer 2 QoS priority of Wi-Fi, and the emergency information element indicating whether the access point supports emergency service over Wi-Fi. These two defined information elements in (IEEE 802.11u, 2011) are not used in the Hotspot 2.0 release 1 (Wi-Fi Alliance Hotspot 2.0, 2012).

Security is important for Wi-Fi access also for the classification in trusted and untrusted accesses to enable the appropriate access towards the 3GPP´s EPC. To enable the end-user devices to securely and more easily and automatically connect to a Wi-Fi access the Hotspot 2.0 specification (Wi-Fi Alliance Hotspot 2.0, 2012) uses several security mechanisms defined in the IEEE 802.11i standard, which is integrated in the IEEE 802.11-2007 standard (IEEE 802.11-2007, 2007). All Passpoint certified devices have to support Wi-Fi Protected Access 2 (WPA2)-Enterprise security to authenticate and secure the air link between the end-user device and the access point. WPA2 uses a four-way handshake and for the encryption on the radio interface the Advanced Encryption Standard (AES). These mechanisms offer a security standard that is comparable to those used in cellular networks. All authentication processes used in Hotspot 2.0 use EAP. The EAP-Tunnelled Transport Layer Security (TTLS) and EAP-Transport Layer Security (TLS) authentication

mechanisms are both intended to be used by pure Wi-Fi operators, because they usually do not have a Subscriber Identity Module (SIM)-based authentication infrastructure. Whereas the EAP-SIM and EAP- AKA (using Universal Subscriber Identity Module (USIM)) are intended to be used by Mobile Network Operators (MNOs) which already have set up and running the necessary infrastructure in their cellular networks. Beside the security mechanisms defined in the IEEE 802.11i standard, Hotspot 2.0 uses layer 2 traffic inspection and filtering, where the firewall function residing either in the access point or in a separate device, connected to the access point. Table 6 summarises the security mechanisms supported by Hotspot 2.0.

Table 6: Supported security standard protocols by the Hotspot 2.0

Credential Type	RFC	EAP Method	Description
Username/ Password	5281	EAP-TTLS	Username and password with server-side use of certificates. Appropriate for a fixed operator without Home Location Register (HLR) /HSS credential capability.
X.509 Certificate	5216	EAP-TLS	Public key infrastructure based client and server certificates. Appropriate for a fixed operator without HLR/HSS credential capability.
SIM card credentials	4186	EAP-SIM	SIM card credentials and 2G network cryptographic algorithms for the challenge response.
USIM	4187	EAP-AKA	USIM/SIM card credentials and 3G network cryptographic algorithms for the challenge response. This method is mandatory within the 3GPP architecture for the S2a interface.

At the time of writing this thesis, the Hotspot Release 2 is being drafted. The main changes would be the addition of operator controlled policies. While the Hotspot 2.0 Release 1 provide the end-user devices with information about available networks it is still the end-user device which makes the decision based on the information about the available networks, the pre-configured operator policies. The Hotspot Release 2 specification will add the capability for the distribution of operator-specific policies to the end-user device. This is a paradigm shift from terminal-based control to network-based control. The ANDSF provides already a network-controlled, terminal-assisted control mode with the Point of Decision at the terminal. The ability to provide policies through the already existing ANDSF and through the still to be defined Hotspot 2.0 Release 2 access point may cause problems:

- If two different kinds of policies are sent to the end-user devices, the 3GPP policies by the ANDSF and non-3GPP providers by a Hotspot 2.0 access point, conflicts can occur.
- The content of 3GPP managed objects and the content of the Wi-Fi Alliance managed objects impacting the WLAN network selection procedures can be inconsistent.

To prevent and resolve these conflicts and inconsistencies caused by the simultaneous use of ANDSF and Hotspot 2.0 Release 2 policies 3GPP is currently working on the technical report (3GPP TR 23.865 V12.1.0, 2013) evaluating 3GPP (ANDSF) and Hotspot 2.0 access network selection procedures. The findings of the technical report could result in changes to current specifications.

3.4.3 Media Independent Handover Services

To provide inter-system mobility in HetNets 3GPP defined the ANDSF, the Wi-Fi Alliance defined the Hotspot 2.0 specification, and IEEE proposes a solution called Media Independent Handover Services (MIHS) defined in (IEEE 802.21, 2009).

The IEEE 802.21 standard (IEEE 802.21, 2009) aims to optimise handovers between HetNets that include both wireless and wired media. Therefore, link-layer intelligence and network environment information are provided to upper layers independent of the Mobile Node (MN) and access network technologies. To get as much information as possible about the surrounding networks, the MN and the network infrastructure are used as information sources. The minimal requirement for processing an inter-system handover is that the MN is a multi-modal device. A multi modal device has at least two interfaces. However, if both interfaces are wireless, severe interference problems can arise when they are simultaneously active. To avoid interference and also to save battery, the interfaces are often used sequentially. This also aligns with the IEEE 802.21 standard framework media independent optimisation (IEEE 802.21, 2009) by providing a generic interface to the upper layer mobility protocols. Within the IEEE 802.21 framework, no handover decisions are made. Even so, the framework provides a generic interface; the media specific technologies have to enhance Service Access Points (SAPs) and primitives to fulfil the generic abstraction. The Media Independent Handover Function (MIHF) is a logical entity and it is the interface

between the media specific technology and the MIH users, the layer 3 or higher mobility protocols. The MIHF provides the following services to higher layers:

- Media Independent Event Service (MIES).
- Media Independent Command Service (MICS).
- Media Independent Information Service (MIIS).

In order to benefit from the MIES, MIH users, as well as MIHFs have to subscribe to events of interest to receive these specific event notifications. The events of interest may include specific state transitions or link layer changes, for example the state of a link layer changing to either up or down, notifications about handover completion, and reports of changes in link conditions that have exceeded a specific threshold. But events can also be predictive; for example, a decrease of the signal strength in a wireless access environment can be an indication that the link layer connection will be lost in the near future. These events facilitate handover decision for the upper layers and can optimise the whole handover process by improving the co-operation of the link layer and upper layer mobility processes.

MICS commands are sent from MIH users to the MIHF, while the interfaces receive commands from the MIHF. The receipt of commands at the MIHF can cause event indications to notify subscribed MIH users of a forthcoming event, such as a handover or a link layer change. The command service enables the MIH user to get dynamic information about the actual situation on the link layer, such as Signal-to-Noise Ratio (SNR) and the Bit Error Rate (BER). In addition, beside other possible applications, commands are used to subscribe or unsubscribe to/from events, configure thresholds for report events, activate actions on the link layer and even link layer resource reservation is possible with MICS. Even though all these possible applications with MICS are defined in the IEEE 802.21 standard this does not imply that all the possibilities with MICS can be exploited in every case, because it depends on the support of the access network technologies and the extension of the supported commands.

The MIIS includes a generic mechanism to allow providers and mobile users to exchange information on possible handover access network candidates. This information is mostly of a static nature contrary to the MICS. The MIIS provides the

information through Information Elements (IE) which can be classified in three groups:

- The general information and access specific information. It provides an overview of the available networks within an area associated with information such as the cost, QoS on the link layer, used frequency bands, the maximum data rate of the link layer, the operator of the network, or the roaming partners.
- Point of Attachment (PoA) specific information. PoAs are network entities which terminate the layer 2, such as an access point (Wi-Fi) or a base station (WiMAX). The IEs belonging to this group provide information about the PoAs of the available networks including but not limited to the channel range, the link layer address and the geographic location of the PoA.
- Access network-, service- or vendor specific IEs. Such elements may provide network information about the supported higher layer services on the supported PDNs.

The main purpose of the mechanisms defined in the IEEE 802.21 standard is to speed up the handover and allow the decision making entity to select the most suitable and appropriate access network to handover to.

3.4.4 Wi-Fi Cell Change

The change of an end-user device from one Wi-Fi cell to another Wi-Fi cell is called roaming. Whenever an end-user device is associated with an Access Point (AP) and changes the AP the term roaming is used. The roaming process can last several 100 ms and therefore roaming is rather slow on legacy Wi-Fi and not suitable for real-time traffic. The IEEE 802.11r standard (IEEE 802.11r, 2008) defines the fast roaming by introducing a concept where the initial handshake with the new target AP is performed before the end-user device roams to the target AP. This kind of roaming is called fast transition. The IEEE 802.11r standard reduces the security overhead of the IEEE 802.11i standard without introducing new security vulnerabilities. The other focus of the standard is the resource reservation at the target AP. The security and QoS states are established before the transition to the new AP takes place and therefore the time consuming processes can be removed from the time-critical re-association process.

There are two protocols defined in the IEEE 802.11r standard:

- **Fast transition protocol:** This is executed if the end-user device makes a transition to a target AP and does not require a resource request prior to its transition.
- **Fast transition resource request protocol:** This is executed if the end-user device requires a resource request prior to its transition. It is optional.

If a STA moves from its current to the target AP utilising the fast transition protocols, the messages exchange is performed by one of the two methods:

- **Over the Air**: The messages are sent from the end-user device directly to the target AP.
- **Over the Distribution System (DS):** The messages are relayed between the end-user device and the target AP by the current AP.

The IEEE 802.11r standard enables Wi-Fi to support real-time data, because the target roaming time is less than 50 ms which would not result in any perceptible audio and video degradation.

3.4.5 Gathering WLAN Key Performance Indicators

The IEEE 802.11k standard (IEEE 802.11k, 2008) is an amendment to (IEEE 802.11-2007, 2007) for Radio Resource Management (RRM). The 802.11k standard is integrated in (IEEE 802.11-2012, 2012). The IEEE 802.11k standard provides the capability to a STA or AP to command a STA to perform measurements. The results of the measurements can be provided to other STAs, APs, to AP controllers or to upper layers in the communication stack. Measurements can be done on the physical layer to indicate the various signal values or the channel load or on the MAC layer to indicate the values of the various counters such as numbers of retransmission, number of successfully sent frames etc. In the following the request/report measurements are listed:

- **Beacon:** The beacon request/report pair enables a STA/AP to request a list of reachable APs on specific channel(s) from a STA.
- **Measurement Pilot:** The measurement pilot frame is an action frame containing a subset of a beacon frame which is transmitted periodically by an

AP to assist a STA with scanning. The measurement pilot frame contains only a single report frame.

- **Frame:** The frame request/report pair is used to get the number of all received frames, the average power level for these frames, and the BSSID of the transmitter.
- **Channel load:** The channel load request/report pair is used to get the channel utilisation as measured by the STA.
- **Noise histogram:** The noise histogram request/report is used to get a power histogram of non-IEEE 802.11 noise power. The measurements are executed when the channel is idle and the STA is neither transmitting nor receiving a frame.
- **STA statistics:** The STA statistics request/report pair is used to get various values from STA counters and BSS average access delay.
- **Location:** The location request/report pair is used to get the requested location as latitude, longitude, and altitude. The requested location can be either the location of the requestor (AP) or the location of the reporting STA.
- **Measurement pause:** The measurement pause request is a single message with no response message. It is used to pause a measurement for a certain time.
- **Neighbour report:** The neighbour request is sent from a STA to an AP to get information about neighbour APs that are possible candidates for roaming. The neighbour report is sent as an answer from the AP to the STA, containing the requested information.
- **Link measurement:** The link measurement request/report is used to get the Radio Frequency (RF) characteristics of a STA to STA link. It indicates the instantaneous quality of the link.
- **Transmit stream/category measurement:** The transmit stream/category request/report pair is used to enable a QoS STA to query another peer QoS STA about the condition of an ongoing traffic stream link between them.

These gained WLAN KPIs are very important and valuable for traffic steering mechanisms. Before the IEEE 802.11k standard was specified the decision on WLAN roaming was solely based on the Received Signal Strength (RSS). But a very good RSS does not guarantee a high throughput. On the contrary, if a the RSS is high, all the

surrounding STAs connect with this specific AP which will lead to an overutilisation of this specific AP and therefore this leads to a massive degradation of the throughput while nearby APs with lower RSS but much higher spare capacity are underutilised.

The IEEE 802.11k standard supports an interface to provide the measurement results to upper layers. The measurement results are stored in the extended part of the Management Information Base (MIB), which is specified by the IEEE 802.11k standard. These values in the extended MIB can be requested from upper layers in the protocol stack through the Simple Network Management Protocol (SNMP) using GET queries with the specification of the Object Identifier (OID).

Upper layers can initiate radio measurements on a STA by performing a MIB.SET operation on the RRM MIB. Each STA maintains a single `dot11RRMRequestTable1` in the MIB used to initiate RM measurement requests. The radio measurement results of one measurement request are spread across multiple RRMReport tables. But all results belonging to one measurement can be identified by an `xxxRprtRqstToken`. Therefore, the results of a measurement can be collected by searching the appropriate `xxxReportTables` and retrieving all reports with the matching request token.

3.4.6 Mobility-Based Strategies for Traffic Steering in Heterogeneous Networks

In the journal paper (Munoz et al., 2013) different mechanisms of traffic steering in HetNets are introduced. First a survey of traffic steering techniques in idle mode is presented. The idle mode is the state in which no dedicated resources have been established for the UE in the RAN. The idle mode cell reselection algorithm according to the 3GPP standard is described. The same analysis is then applied to the connected mode. The connected mode is the state in which dedicated resources have been established for the UE. Furthermore, a fuzzy-logic-based algorithm that optimises network parameters for traffic steering is proposed.

In idle mode, the cell to which the end-user device is attached is called the camped cell. The selection of a cell when the UE is powered on is called cell selection. If a cell selection is made due to mobility of the UE it is called cell reselection. Four traffic steering techniques are briefly introduced, whereas a specific traffic steering

technique using absolute priorities is discussed as the main traffic steering technique in the paper (Munoz et al., 2013).

Hierarchical cell structure can be used to provide different priorities to large and small cells for cell reselection. This can be useful for example to direct UEs at a high speed into a large cell and UEs at a low speed into small cells.

Cell barring can be used to bar a cell at which the UE cannot camp.

Basic biasing is the principle that is applied to the CRE mechanisms previously discussed in section 3.3. Offsets are applied to cell reselection related parameters to make the cell more or less attractive for the cell reselection process. LTE and LTE advanced allow the use of basic biasing only to cells belonging to the same RAT. This clearly limits the possibilities of the traffic steering mechanisms.

Absolute priorities are a set of parameters applied in the context of different frequencies and RATs that can be used to steer end-user devices towards a different RAT or frequency. Different priorities are allocated to network layers and thus the end-user device distribution can be steered in idle mode by changing the priorities of the layers. Absolute Priorities allows the operator to prioritise inter-frequency and inter-RAT network layers during the cell reselection process. If there are multiple cells of different priorities suitable for cell reselection, cells of higher priority will take precedence over lower priority frequency and RAT. The cell reselection is based on absolute priorities and on threshold definitions that are exceeded or undercutted.

The possibility and the impact of adjusting the absolute priorities for traffic steering purposes are discussed in (Munoz et al., 2013). The allocation of absolute priorities to network layers is limited according to (3GPP TS 36.304 V11.6.0, 2013).

- Layers with the same RAT and frequency must have the same absolute priorities despite different cell sizes.
- Different RATs sharing the same absolute priorities are not allowed.

These limits result in less flexibility of allocating absolute priorities to network layers. For example, a High Speed Packet Data Access (HSPA) macro and picocell operating on the same frequency must have the same absolute priorities allocated, or different RATs such as LTE and HSPA may not have the same absolute priorities allocated. In

(Munoz et al., 2013) it is shown that with the allocation of different absolute priorities to network layers the distribution of idle mode UEs can be effectively modified.

In connected mode traffic steering can be performed by forced handovers, cell barring, or adjusting handover parameters. In (Munoz et al., 2013) the handover parameters are adjusted to satisfy a specific traffic steering policy. The adjustment of handover parameters is the simplest mechanism because it does not need to be user-specific like the traffic steering with forced handovers, and it is not as rigid as cell barring. The paper focuses on adjusting inter-RAT handover parameters corresponding to the 3GPP defined B2/A3 events to trigger handover either from LTE to HSPA or from HSPA to LTE. Handover parameters are typically adjusted to avoid low signal levels or quality. In the context of HetNets there are many other reasons to trigger a handover. 3GPP allows using different thresholds for the same handover triggering event according to particular policies. For example, radio-driven handovers are processed if the signal level is poor to avoid a loss of the connection, while traffic steering-driven handovers are triggered even when the signal level is good, but e.g. the throughput can be maximised with a handover to another RAT. In the paper (Munoz et al., 2013) it is shown that with an appropriate setting of the thresholds for radio-driven and traffic steering-driven handover the small cells can be also extended quite similar to the CRE with LTE-Advanced cells. Finally a fuzzy-logic-based Self-Organising Network (SON) algorithm for traffic steering by adjusting handover parameters is proposed in the paper. The simulation results show that the proposed algorithm can automatically adapt to context variations of the end-user spatial distribution and more picocells are used for offload traffic.

3.4.7 Mobility Enhancements for LTE-Advanced HetNets

In the journal paper (Pedersen et al., 2013) a state-of-the-art HetNet scenario with macro and small cells deployed on different carriers, while using inter-site carrier aggregation is presented. Furthermore, HetNet mobility enhancements are proposed. The paper considers only the connected mode state of Radio Resource Control (RRC), but not the idle mode state.

Carrier aggregation is a method of aggregating multiple carriers either contiguous or non-contiguous for an end-user device. Carrier aggregation was introduced in Release 10 for LTE-Advanced. The resulting increase of bandwidth leads to an increase of the

data rate. Inter-site carrier aggregation is targeted in 3GPP Release 12, which has not been finalised by the time of writing this thesis. To aggregate carriers of different sites the UE has to be able to handle multiple connectivities to different eNBs. The ability of the multiple connectivity mode has also been introduced in the as yet not finalised 3GPP Release 12. The ability of dual connectivity enables UE to, for example, split the user plane and the control plane towards different cells. In the paper the following scenario is used: the control plane is sent through the macro cell whereas the user plane is sent through the small cell. Macro and small cells use different frequencies. UEs always have a downlink connection from the macro cell which is called Primary Cell (PCell). If the UE is in the range of a small cell it can have this particular cell configured as the Secondary Cell (SCell) and thus benefit from inter-site carrier aggregation to achieve higher data rate due to the higher accessible bandwidth. Macro and small cells are assumed to be connected via either the X2 interface or fibres using other protocols. Assuming that there are 2 overlapping macro cells containing 1 and 2 small cells. A UE moves through the first macro cell containing 1 small cell. When the UE is entering the coverage area of the small cell this cell is added as an SCell and when exiting the coverage area of the small cell the SCell is removed. Arriving at the cell edge of the macro cell a PCell handover to the other macro cell is performed which contains 2 overlapping small cells. When the UE enters the coverage area of the first small cell this cell is added as the SCell. When the UE enters the coverage area of the second small cell while the first cell is the SCell an SCell change is performed and when the UE exits the coverage area of the second SCell the SCell is removed. While the UE has a connection with an SCell the PCell connection remains. LTE uses a UE-assisted network-controlled mobility paradigm in RRC connected mode. This means that when using current LTE-Advanced carrier aggregation mobility functions the network will have to send an RRC message to the UE whenever an SCell is added, changed or removed. These network actions are triggered by measurement reports sent from the UE to the network via uplink RRC signalling. In an area with a dense small cell deployment the RRC signalling from SCell operations can be significant compared to managing PCell mobility. The proposed approach in the paper (Pedersen et al., 2013) is that the SCell management is done using UE autonomous decisions with a certain degree of network control.

The PCell mobility management is performed the current way: UE-assisted and network-controlled with the Point of Decision located at the terminal. But for the SCell mobility management a UE autonomous solution is investigated. Therefore, the UEs are configured with a list of candidate cells sent via a dedicated signalling channel or via a broadcast channel from the network to the UE(s). The list would include candidate SCells and their system information and Random Access Channel (RACH) preamble to use. On the network side the candidate SCells have to be informed about the UEs eventually using the cell as an SCell. The candidate SCells have to be aware of the UE identity and the UE´s PCell. When the UE detects a candidate SCell and this SCell fulfils certain criteria, it can autonomously request this cell as an SCell. This process is valid for the adding and changing of an SCell. The trigger for this process is based on the measurement results done by the UE itself, for example the signal strength of a candidate SCell exceeds a certain threshold (adding SCell) or that the signal strength from another SCell is better than the current SCell (changing SCell). After the SCell received the request from a UE over the RACH the SCell informs the PCell that it is now the serving SCell. Afterwards, the data traffic is sent from the SCell to the UE. When a UE leaves the coverage area of an SCell the signalling is sent to the PCell to inform the PCell that the SCell is no longer the current SCell (remove SCell). The number of messages on the radio link is reduced with the UE autonomous SCell mobility management compared with the UE-assisted network-controlled SCell mobility management. Addition of SCell is performed faster than traditional SCell addition because the SCell as well as the UEs are preconfigured and informed. These findings are approved by simulation results in the paper.

3.4.8 Bringing Always Best Connectivity Vision a Step Closer

In the journal paper (Louta and Bellavist, 2013) the challenges and perspectives that the ABC principle raises, are identified and discussed. To enable the ABC to the end-user devices, they always have to have a connection towards an appropriate RAN; every time, independent of the mobility. The Access Network Selection (ANS) is very important to provide always best connectivity to the end-user device. In the paper (Louta and Bellavist, 2013) the critical aspects and research challenges of ANS are identified and discussed.

The complexity of future communication systems will increase because of the heterogeneity of the access networks with diverse features and various network technologies, all interworking with one and another to provide an appropriate service for the end-user in a cost efficient way. The ABC principle falls within the realm of handover management procedures, including seamless intra-system and inter-system handovers. A handover consists of three phases: the handover initiation phase, the handover decision phase and the handover execution phase. The ABS principle is essentially implemented in the handover decision phase. In (Kassar, Kervella and Pujolle, 2008) it has been shown that traditional handover decisions based on RSS work for homogeneous networks but are not sufficient for HetNets. Besides using the RSS to make a handover decision, additional criteria such as user requirements and preferences, terminal/service/application characteristics and capabilities, network conditions, costs, and security-related aspects should be considered. An overview of the ANS standardised efforts is given in the paper (Louta and Bellavist, 2013). These are the IEEE 802.21 standard (IEEE 802.21, 2009) (discussed in more detail in this thesis in section 3.4.3), the ANDSF defined by 3GPP (3GPP TS 23.402 V12.3.0, 2013) (3GPP TS 24.312 V12.3.0, 2013) (discussed in more detail in this thesis in section 3.4.1), and the IEEE 1900.4 standard. The IEEE 1900.4 (IEEE 1900.4, 2009) standard defines the architectural building blocks enabling network-device distributed decision making for optimised radio resource usage in heterogeneous wireless access networks. The aim of the standard is to improve overall composite capacity and QoS of wireless systems in a multiple RAT environment, by defining suitable system architecture and protocols (IEEE 1900.4.1, 2013) that will facilitate the optimisation of radio resource usage. This is done by exploiting information exchanged between network and mobile terminals. The amendment 1 (IEEE 1900.4a, 2011) addresses the coexistence of secondary systems and spectrum management in white space frequency bands.

The ANS procedure can be either network-controlled or terminal-controlled. The network-controlled ANS procedure is typically controlled by a network-operator-related intelligent entity, residing in the network operator´s domain. The terminal-controlled ANS procedure is divided into two categories in the paper (Louta and Bellavist, 2013). One is the terminal-controlled network-assisted ANS procedure with a terminal-related entity undertaking the ANS task by exploiting network-related

information. The other one is the network-controlled terminal-assisted ANS procedure where a network-related entity considers information and measurements gathered from the terminal to perform the ANS. The paper mentions that subscribers fear that network operators in case of network-controlled and network-controlled terminal-assisted ANS processes are unfairly performed because of the operator´s preferences, requirements, and constraints that can result in conflicting goals and business policies. Terminal-controlled, network-assisted ANS procedures are stated in the paper as more flexible, relieving the network of significant complexity, while they are considered to be an imperative property of 4G ABC environments. The opinion of the paper authors (Louta and Bellavist, 2013) is that coexistence and potential interworking for example by means of a negotiation phase of both types of terminal/network-controlled mechanisms would facilitate the realisation of the ABC vision.

Four ANS initiation reasons are identified in the paper. These are:

- A new service request
- In the case of an active service session
    - If a new wireless access network is identified.
    - If the QoS degrades below a certain threshold.
- Imperative and robustness-related conditions such as current RAT failure, network- or handover execution failure.

The paper lists and discusses ANS decision criteria and identified four groups of criteria which are either static (changes do not often occur, such as user profiles, terminal characteristics etc.) or dynamic (changes do occur often, such as network conditions):

- Link quality (RSS, carrier-to-interference ratio, signal-to-interference ratio, signal-to-noise-and-interference ratio)
- Network availability (coverage, bandwidth availability, call blocking probability)
- QoS-related aspects (considered throughput, delay, jitter, latency, bit error rate, packet loss ratio, average number of retransmission per packet)
- Network reliability (call dropping and handover failures)

There exist several methodologies and algorithms to perform ANS, but it is hard to find classifications of algorithms. There are a few classes of methodologies and algorithms that are outlined in the paper. These are the multi-criteria decision making algorithms, fuzzy-based mechanisms and policy-based methodologies. It is stated in the paper that policy-based methodologies are claimed to be sufficient for handling complexities in 4G systems.

Evaluation methods for ANS are very different for diverse research works because of different input parameters, decision criteria, methodologies, and performance evaluation metrics. That makes differentiation between the various evaluation methods very difficult. The most commonly used performance metrics used for evaluation purposes contain the number of handovers performed, handover success and failure rate, delay associated with the three handover phases (initiation, decision, and execution), and packet loss. In most cases performance evaluation is done with simulations using different simulators.

The authors of the paper propose to add an additional handover phase, the negotiation phase. This negotiation phase may be used to negotiate between the network and the users in order to guarantee successful handovers as much as possible. The negotiation phase should be not mandatory because if a handover is imperative, such as in time-critical situations, the adding of a negotiation phase may lead to a failed handover or high QoS degradation for example resulting in call drops. A negotiation phase could be added if the handover is not imperative to ensure that a user will be admitted to a new network node on the basis of its current load.

The ANS mechanisms should have a certain robustness and ability to handle large numbers of decision criteria, an increasing number of RATs, possibly imprecise data and/or partial knowledge, and an uncertain and highly dynamic environment. Policy-based solutions in combination with machine learning techniques seem to constitute good candidates for a viable answer to the robustness challenge.

The author´s vision of getting closer to always best connectivity is based on cognitive networks to efficiently manage the increasing complexity and heterogeneity by the ability of thinking, learning, remembering, and adapting to changing conditions. A generic architectural framework of a cognitive system is proposed containing the following modules: sensing, reasoning, learning, decision, act, and policy. The system

is continuously sensed by the sensing module that aggregates, correlates, and filters data until a condition that should be further analysed, is identified. The sensing module identifies available access networks, measures and aggregates QoS-related information, observes current context such as terminal velocity, location, and battery status and provides this data to the reasoning module. Data from the sensing module is processed and analysed by the reasoning module and further provided to the learning module. The reasoning module identifies potential actions to be taken and generates a candidate network list by interworking with the learning and policy modules. The reasoning module uses policies to eliminate inappropriate access networks from the candidate network list and it uses data from the sensing module to exclude for example inappropriate cell sizes because of velocity. The learning module is also involved in the process of generating the candidate network list. The decision module selects the action to be taken by interworking with the learning module. Finally, the act module executes the selected action.

Two conflicting classes of cognitive entities have been identified in the paper: the user-related cognitive entity, acting on behalf of the user, and the network-related entity, acting on behalf of the network operator. Both kind of entity are comprised in all the proposed modules and may be implemented in the 3GPP´s ANDSF and/or in the IEEE´s 802.21 MIH entities. But these two cognitive entities have different and possibly contradicting objectives because they belong to different owners and want to achieve as many objectives as possible. But these two cognitive entities have to be able to cooperate with each other in order to achieve a joint ANS decision. Therefore, a careful design of a negotiation process could serve for successful handovers and personalisation by involving the network operator and the user side.

3.4.9 3GPP Study on WLAN – 3GPP Radio Interworking

The Technical Report (3GPP TR 37.834 V12.0.0, 2013) consists of the output on the 3GPP study on WLAN – 3GPP radio interworking. The interworking and integration of WLAN is currently supported by 3GPP specifications at the core network level and seamless and non-seamless mobility. But there are several unsolved issues in 3GPP Release 11, such as that operator deployed WLANs are often under-utilised, if a user connects to an overloaded WLAN the experience is suboptimal, and the unnecessary scanning for WLANs causes an increase of battery consumption. Therefore, the 3GPP

agreed to study potential RAN level enhancements for WLAN – 3GPP interworking in Release 12. The study is divided in two phases:

- Phase 1: Identifies the requirements for RAN level interworking and clarifies the scenarios to be considered in the study while taking into account existing standard mechanisms.
- Phase 2: Includes the evaluation of the benefits and impacts of identified mechanisms over existing functionality including network-based WLAN interworking mechanisms such as the ANDSF. Furthermore, phase 2 identifies solutions addressing the requirements which are identified in phase 1 which cannot be solved using existing standardised mechanisms consisting of:
    - Solutions that enabling WLAN to be included in the operator´s cellular RRM and enable an enhancement of the operator´s control over WLAN interworking.
    - Solutions that enhance the access network selection and mobility taking into account cellular and WLAN access information such as backhaul quality, radio link quality per UE, utilisation, etc.

Both phases are addressed in the Technical Report (3GPP TR 37.834 V12.0.0, 2013). The definition of the two phases indicates clearly that 3GPP aims to integrate WLAN further into the cellular networks and it indicates as well that the operator´s control over WLAN will be increased, which implies a paradigm shift from terminal-controlled (which is the case in nowadays WLAN deployments) to an enhanced operator- or network-controlled and terminal-assisted mode.

The requirements of the phase 1 are the following:

- Bi-directional load balancing between WLAN and 3GPP RANs.
- Solutions should improve performance (no performance degradation through solutions)
- Solutions should improve the utilisation of WLAN when available and not congested. According to the before mentioned requirement: no performance degradation.
- Solutions should reduce or maintain battery consumption.

- Solutions should be compatible with all existing core network WLAN related functionality, e.g. seamless and non-seamless offload, trusted and non-trusted access, MAPCON and IFOM.
- Retain backward compatibility with 3GPP and WLAN specification.
- Solutions should rely on existing WLAN functionality and should avoid changes to IEEE and WFA specifications.
- Per target WLAN system distinction (e.g. based on SSID) should be possible.
- Per-UE control for traffic steering should be possible.
- Solutions should ensure that access selection decisions should not lead to ping-ponging between Universal Terrestrial Radio Access Network (UTRAN)/E-UTRAN and WLAN.

The Technical Report (3GPP TR 37.834 V12.0.0, 2013) proposes 3 solutions for WLAN – 3GPP radio interworking. The first two solutions are network-controlled, terminal-assisted approaches with the terminal as the Point of Decision for traffic steering issues. They both use the UTRAN/E-UTRAN as a source of information providing information to the UE. The UE processes all kind of available information and decides on the traffic steering procedure. The third solution is a network-controlled terminal-assisted approach with the PoD at the network for traffic steering issues. For all 3 solutions the user preferences always take priority over RAN-based or ANDSF-based rules. For example, if a non-operator WLAN is preferred or WLAN is switched off at the terminal.

Solution 1 of the Technical Report (3GPP TR 37.834 V12.0.0, 2013) implies the provision of UTRAN/E-UTRAN assistant information to the UE via broadcast or optionally via dedicated signalling. This solution requires the deployment of the ANDSF. The UTRAN/E-UTRAN assistant information may contain the following parameters:

- Load information indicating in a direct or indirect way the load of UMTS/LTE for example in percentage, in load levels (low, medium, high) or offload preference indicator.
- Resource allocation indicating the maximum resource allocation the UE may receive on UMTS/LTE.

- WLAN thresholds containing the Received Signal Strength Indication (RSSI) threshold, the WLAN BSS load threshold, and WLAN WAN metric (e.g. uplink/downlink estimation of speeds, loading, link status, and whether the WLAN is at capacity) threshold from the Hotspot 2.0 (Wi-Fi Alliance Hotspot 2.0, 2012). The WAN metric from Hotspot 2.0 ANQP element provides information about the WAN link connecting the hotspot to the internet.
- RAN thresholds including the Reference Signal Received Power (RSRP) used with LTE and the Received Signal Code Power (RSCP) used with UMTS.

The policies provided to the UE are enhanced by having the UTRAN/E-UTRAN assistance information. Therefore, the ANDSF Managed Object (MO) defined in (3GPP TS 24.312 V12.3.0, 2013) has to be extended as indicated in (3GPP TR 37.834 V12.0.0, 2013).

Figure 22 shows the overview of the process of solution 1. For a better overview the ANQP requests and the probe requests in uplink direction from UE to the WLAN AP are not shown. The order of the depicted messages is irrelevant. The UE processes all information such as the UTRAN/E-UTRAN assistant information from the UTRAN node Radio Network Controller (RNC) or from the E-UTRAN node eNB, the UE measurements and information provided by the WLAN, the policies provided by the ANDSF or OMA DM mechanisms, and the pre-configured policies to steer traffic IP flows to the WLAN or to the UTRAN/E-UTRAN. Thereby, the UE may replace the thresholds and values of the ANDSF MO with the thresholds and values of the UTRAN/E-UTRAN and WLAN if available.

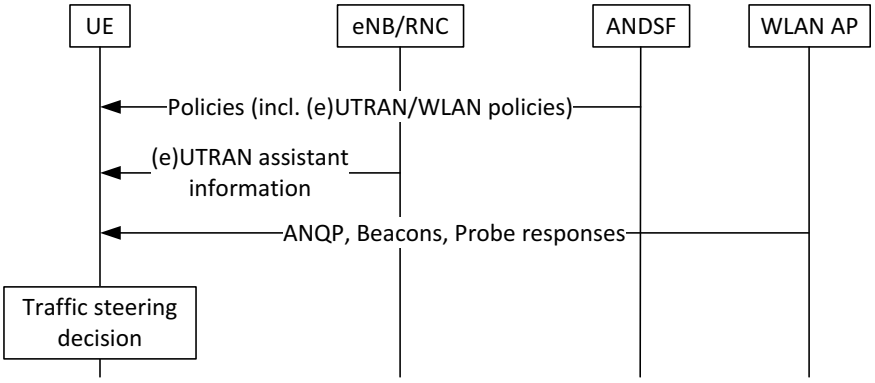


Figure 22: Overview of the process of solution 1

Solution 2 of the Technical Report (3GPP TR 37.834 V12.0.0, 2013) implies the provisioning of offload rules and thresholds by the UTRAN/E-UTRAN through

broadcast and/or broadcast signalling to the UE. The offload rules are specified in the UTRAN/E-UTRAN. The ANDSF may or may not be deployed in solution 2. If the ANDSF is deployed, the ANDSF restrictions concerning certain RANs restrictions (e.g. forbidden or restricted RANs) are valid at any time. But the UTRAN/E-UTRAN rules can restrict access network availability for example if the ANDSF allows two RANs, the UTRAN/E-UTRAN rules may indicate any of the two RANs as not available, even if the one indicated as not available has the higher ANDSF precedence. If the ANDSF is not deployed in the network, the UTRAN/E-UTRAN rules are fully valid as a single source of rules. As in the previously described solution 1 the order of the depicted messages is irrelevant and all the received information at the UE is processed by the UE and the traffic steering decision is made by the UE.

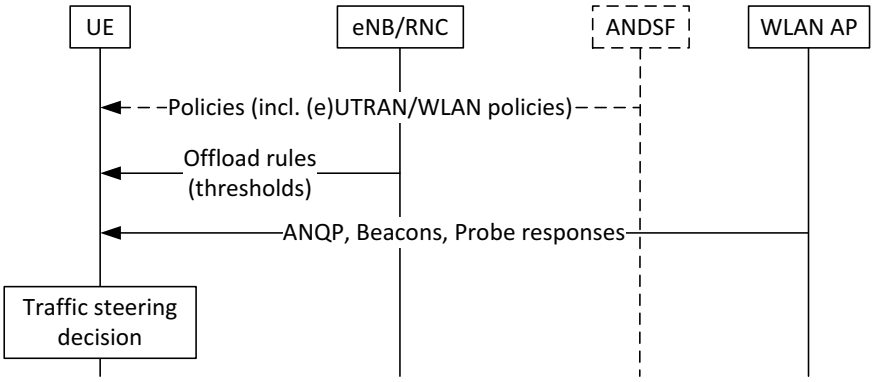


Figure 23: Overview of the process of solution 2

Solution 3 of the Technical Report (3GPP TR 37.834 V12.0.0, 2013) is a network-controlled terminal-assisted solution for traffic steering where the PoD for traffic steering issues is in the network. The decision is provided to the UE with traffic steering commands. The ANDSF may or may not be deployed. Only dedicated signalling is used in this solution. The dedicated traffic steering commands are able to override the order of ANDSF precedence of RANs. But restricted or forbidden RANs by the ANDSF cannot be used with dedicated traffic steering commands.

The network controls the measurement process by configuring the UE measurement procedure. The configuration of the UE measurement process is done by providing information about the candidate WLAN to be measured by the UE and by providing measurement events to the UE. The information about the candidate WLAN contain the BSSID, the SSID, the HESSID, the Domain name list containing names of the entity operating the WLAN access network, and the operating class channel number

indicating the candidate WLAN frequency. The measurement events may contain the following structures:

WLAN becomes better or worse than a threshold. 3GPP Cell's radio quality becomes worse than threshold1 and WLAN's radio quality becomes better than threshold2. WLAN's radio quality becomes worse than threshold1 and 3GPP Cell's radio quality becomes better than threshold2.

The measurement report sent from the UE to the UTRAN/E-UTRAN may contain the following information:

The Received Channel Power Indicator (RCPI) indicating the Radio Frequency (RF) power in the selected channel for a received frame, the Received Signal to Noise Indicator (RSNI) indicating the signal to noise plus interference ratio of a received frame, the BSS load containing the current STA population and traffic levels in the BSS, and the WAN metrics.

Based on this measurement report the UTRAN/E-UTRAN makes a decision on traffic steering and sends the command to the UE where the traffic steering is executed according to the traffic steering command. An overview of the process of the traffic steering solution 3 is provided in Figure 24.

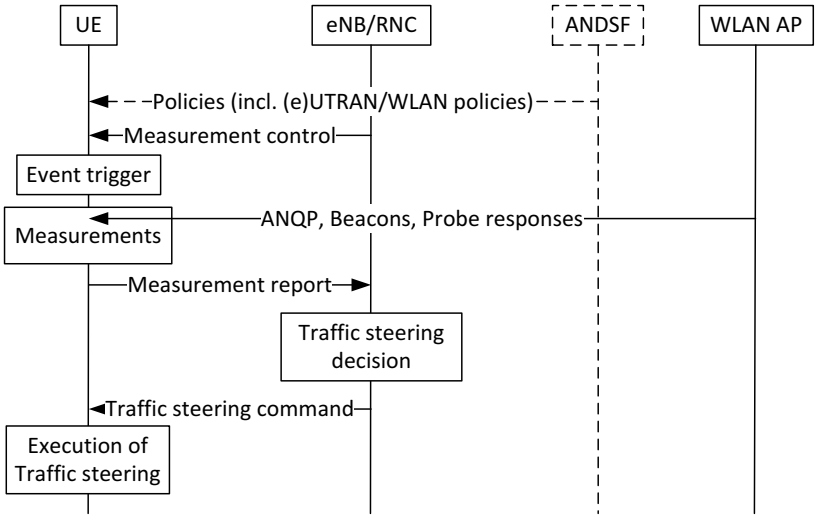


Figure 24: Overview of the process of solution 3

3.5 Future Trends of Mobile Networks

The METIS project, an EU 5G project with the aim of laying the foundation for 5G systems and building consensus prior to standardisation, is presented in (Osseiran et

al., 2014). In the following the main technical relevant challenges and directions are listed, which are identified in (Osseiran et al., 2014).

- Traffic will continue to increase (estimation of a 1000x by 2020 compared with 2010).
- Amount of connected devices will increase dramatically. 50 billion connected devices are estimated by 2020. This means a 10-100 times higher number of connected devices.
- 10-100 times higher user data rate. Supporting data rates of up to 5 Gb/s.
- New devices will significantly contribute to that increase (e.g. probes, sensors, meters, machine-to-machine modules).
- Energy savings. On both sides, the network and end-user devices the energy consumption has to be reduced. 10 times longer battery life for low-power massive machine communication.
- 5 times reduced end-to-end latency, supporting 5 ms end-to-end delay.
- Higher availability and reliability to support safety-critical applications. Provide the possibility to applications to request ultra-reliable links and ultra-low latency.
- Infrastructure densification to increase capacity, increase energy efficiency of radio links, and enable better exploitation of the spectrum.
- Additional spectrum in the higher frequency range >6 GHz. Multi antenna technologies advances such as massive MIMO.
- Improve mobility and load balancing capabilities between different RATs.
- Enable device to device communication.
- Utilise the licensed and unlicensed frequency bands.
- Usage of low, medium, and high dense cells.
- Network Functions Virtualisation (NFV) as a driver for cheaper devices, energy efficiency, reusability, separation of hardware and software.

The above list is not exhaustive.

3.6 Summary

This chapter provides the related work which presents the state-of-the-art in the field where this research is placed. The 3GPP´s EPS architecture is presented with the QoS,

the policy and charging control capabilities. Furthermore, the traffic offloading principles and the various architecture variants are presented which are standardised within the 3GPP. In this research area the small cells and HetNets are analysed which brings up new challenges with interferences between small cells and macro cells. 3GPP provides multiple protection mechanisms to reduce interference between macro- and small cells, which are described in the related work chapter. The already available traffic steering mechanisms from 3GPP, from non-3GPP standards as well as from literature have been studied and described. The future containing the challenges and problems of mobile networks which is targeting the 5G of mobile networks is outlined to round off the research area.

4 Design Selections

In this section the design decisions for the thesis are presented. This has the aim of reducing the many possibilities defined in the standards in order to focus on the main problem carved out in this thesis.

4.1 3GPP Access Technologies

Several generations of 3GPP RATs exist from legacy 3GPP technologies such as GSM, GPRS, UMTS, HSPA to recent technologies such as LTE and LTE-Advanced. In this thesis LTE and LTE-Advanced are selected to be considered as the 3GPP RATs. To cope with the predicted traffic increase, access technologies supporting very high throughput are required. This is the case with LTE and LTE-Advanced, whereas legacy 3GPP RATs support rather less throughput. The legacy 3GPP RATs are still used in present and future networks, but they are only able to serve a fraction of the predicted traffic growth.

4.2 Non-3GPP Radio Access Technologies

Today, the two non-3GPP RATs available are WiMAX and Wi-Fi. The differences between these two RATs is that WiMAX, like the 3GPP RATs, is a network-controlled technology which includes also appropriate QoS solutions such as QoS on a per flow basis and network-based mobility management mechanisms. Besides the support of nomadic and pedestrian mobility of the end-user device the support of a much higher end-user device speed (60-120Km/h) has been introduced in the IEEE 802.16e amendment (IEEE 802.16e, 2006). The IEEE 802.16e amendment has been integrated into the (IEEE 802.16, 2009) standard. With the amendment IEEE 802.16m (IEEE 802.16m, 2011) WiMAX fulfils the International Telecommunication Union Radiocommunication Sector (ITU-R) International Mobile Telecommunications-Advanced (IMT-Advanced) requirements (ITU-R M.2134, 2008) on 4G systems. On the other hand, Wi-Fi is a terminal-controlled technology. QoS mechanisms are also available with the amendment IEEE 802.11e (IEEE 802.11e, 2005) where 4 QoS classes are defined (Best Effort, Video, Voice, and Background). The IEEE 802.11e

standard is integrated in the 802.11-2007 standard (IEEE 802.11-2007, 2007). The QoS adaptations through the 802.11n standard (IEEE 802.11n, 2009) including frame aggregation mechanisms are integrated in the 802.11-2012 standard (IEEE 802.11-2012, 2012), but compared with the WiMAX QoS framework the Wi-Fi QoS support is rather rudimentary, since with these amendments it is still not possible to give any guarantees about bandwidth and delays. For the time being, since the offload of traffic from 3GPP networks to non-3GPP networks commonly consists of flows that are not prone to jitter and delay, QoS is not that important to offload mechanisms. The more important aspect is that mobility is supported in a seamless manner. And this is fulfilled with Wi-Fi access networks; because mobility within Wi-Fi access networks as well as inter-system mobility with 3GPP access networks has been tremendously improved with the introduction of the Hotspot 2.0 specification (Wi-Fi Alliance Hotspot 2.0, 2012). Another aspect that differentiates these two access technologies is that Wi-Fi is deployed more widely than WiMAX. There are several reasons for this, such as that Wi-Fi technology is very cheap to deploy and there is no effort to take with regards to spectrum allocation, because Wi-Fi is operated in unlicensed spectrum, whereas WiMAX is more expensive to deploy and is operated in licensed spectrum.

ABI Research published in September 2013 an article about the development of Wi-Fi (ABI Research, 2013). The following is an excerpt of it.

ABI research forecasts Carrier Wi-Fi access point shipments in 2018 to reach 9.7 million with the Asia-Pacific region accounting for 70% of that number. In the U.S., a very successful model was deployed by Cable Wi-Fi: an alliance formed of five of the biggest cable operators in the country including Bright House Networks, Comcast, Cablevision, Cox, and Time Warner Cable. Customers of any of the alliance members can roam seamlessly in the biggest Wi-Fi network in the U.S. with more than 150 000 hotspots.

Another statement of an ABI Research article (ABI Research, 2013) is that the worldwide carrier Wi-Fi deployments reached a total of 4.9 Million hotspots in 2012 and is anticipated to proliferate and surpass 6.3 million by the end of 2013.

Today, it is very hard to find any actual information about the amount of WiMAX operators, subscribers, countries etc. Years ago, when WiMAX was booming, the WiMAX Forum maintained a global WiMAX distribution map updated every quarter

year, where the mentioned data could be looked up. But this website is no longer maintained. ABI Research published figures of WiMAX compared with LTE (ABI Research, 2012).

In South Korea, the United States, and Japan, the number of LTE subscribers surpassed that of WiMAX subscribers in 4Q11, 1Q12, and 2Q12, respectively. "Japan, South Korea, and the United States used to have strong mobile WiMAX proponents, so while the momentum and future of WiMAX and LTE are clear, it is somewhat surprising to see how long the subscriber crossover has actually taken," says research director Phil Solis. "In mid-2014, even subscribers to LTE in Time Division Duplex (TDD) mode will have surpassed WiMAX subscribers at which point WiMAX subscribers will begin their permanent, slow decline."

Wi-Fi is deployed all over the world due to the fact that it is cheap and there is no need of owning specific spectrum ranges to be able to operate Wi-Fi access networks. This stands in contrast to the WiMAX technology, where there are countries that do not operate WiMAX networks. Furthermore, Wi-Fi is classified nowadays as a trusted access technology and, as a result of this; the access to the 3GPP EPC is facilitated and equal to the EPC access through WiMAX. Therefore, the technology investigated in this thesis for non-3GPP radio access and traffic offloading is Wi-Fi.

4.3 Host-Based Versus Network-Based Mobility Protocol

Two types of mobility protocols exist: the host-based mobility protocols and the network-based mobility protocols. The difference between these two mobility protocols is that mobility is managed by end-user devices when a host-based mobility protocol is used and the network manages mobility when a network-based mobility protocol is used. Therefore, the mobility related signalling is performed either by the host itself, when a host-based mobility protocol is used, or by the network, when a network-based mobility protocol is used. As a result of this, the end-user devices have to support mobility related signalling when a host-based mobility protocol is used and therefore the end-user devices need to be modified. In contrast, with the network-based mobility protocols end-user devices are not aware of any mobility and no modification at the end-user devices is necessary. A host-based mobility protocol used in the EPS on the S2c interface is the DSMIPv6 (IETF RFC 5555, 2009). With DSMIPv6 used on the S2c interface the access to the EPC can be enabled for both

trusted and non-trusted non-3GPP access networks. Network-based mobility protocols used in the EPS are PMIPv6 (IETF RFC 5213, 2008) and GTP (3GPP TS 29.274 V12.3.0, 2013), (3GPP TS 29.281 V11.6.0, 2013) both used on the S2a, S2b and S5/S8 interfaces. But these two network-based mobility protocols are used on the interfaces in an "either-or" principle. Network-based mobility protocols have several advantages over the host-based mobility protocols:

- Network-based mobility protocols reduce the complexity of the end-user devices compared to host-based mobility protocols which add complexity to the end-user devices through extending the IP stack with an implementation of IP mobility signalling and the ability of detecting end-user device movement.
- WiMAX provides with MIP a host-based and with PMIPv6 a network-based mobility protocol. But end-user devices with an extended IP stack that provides the ability for mobility signalling are not widely available and operators prefer mobility control remaining in the network. Therefore, PMIPv6, the network-based mobility protocol, has prevailed.
- If DSMIPv6 is used with a non-trusted non-3GPP access, an IPsec tunnel has to be established between the end-user device and the ePDG. Furthermore, a continuous DSMIPv6 tunnel has to be established between the end-user device and the PDN GW. This results in a tunnel in tunnel solution which is not as efficient as the PMIPv6 or the GTP solution, where no tunnel in tunnel has to be established. The tunnel in tunnel situation is shown in Figure 25.

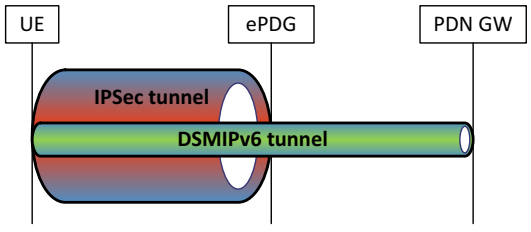


Figure 25: Tunnel in tunnel solution with DSMIPv6 mobility protocol

All these drawbacks of the host-based mobility protocol lead to the decision that network-based mobility protocols are more appropriate and therefore the network-based mobility protocols are the ones to be further considered in this thesis.

4.4 Mobility Protocol Selection

Since the decision in section 4.3 is made for the network-based mobility protocols and against the host-based mobility protocols, the eligible mobility protocols are therefore the network-based ones. The 3GPP definitions of the EPS specify two variants of network-based mobility protocols on the S5/S8 interface between the SGW and the PDN GW, as well as on the S2a interface to enable trusted non-3GPP access networks to connect to the EPC and on the S2b interface to enable non-trusted non-3GPP access networks to connect to the EPC. These two network-based mobility protocol variants are the GTP and the PMIPv6. GTP uses on the control plane an enhanced version of the GTP version defined for the 3GPP 2G and the 3G networks, defined in (3GPP TS 29.274 V12.3.0, 2013). On the user plane, the GTP version deployed in 3G networks (3GPP TS 29.281 V11.6.0, 2013) is used. The PMIPv6 protocol is defined in (IETF RFC 5213, 2008) on the basis of (IETF RFC 3775, 2004).

GTP and PMIPv6 are both network-based mobility protocols, but GTP includes additional functionality besides mobility. Table 7 compares the range of functions supported by both mobility protocols GTP and PMIPv6.

Table 7: Functions supported by GTP / PMIPv6

	GTP	PMIPv6
QoS signalling	+	-
Bearer signalling	+	-
Packet forwarding during handover	+	-
Mobility tracking	+	+, always active
Paging	+	+, always active
Network-based mobility management scheme	+	+

As shown, GTP supports QoS signalling and is able to establish and modify bearers. In case of a handover, data can be forwarded to the new cell since GTP can be used to establish the necessary tunnels for forwarding data on the user plane. Paging is not supported with PMIPv6, but since the end-user device is always active with PMIPv6, paging is not needed. This always active behaviour of the end-user device increases the power consumption. From the PMIPv6 point of view, functions to reduce the power consumption of the end-user device are not supported. The end-user device is always active. The ability to change the state of the end-user device from active to idle or sleep mode is left to the access technology and is therefore not supported by the PMIPv6.

With PMIPv6, one tunnel per UE and per PDN is established, unlike GTP where a maximum of 11 tunnels, each of them realising a bearer, can be established per UE. The number of maximum GTP tunnels is limited to 11 due to the 4 Bit EPS Bearer Identity (EBI) field with 5 reserved values (3GPP TS 24.007 V12.0.0, 2013). GTP can provide a much finer grained QoS than PMIPv6, because every GTP tunnel establishes an EPS bearer and every EPS bearer has an assigned QoS. PMIPv6 can establish only one GRE tunnel per UE and per PDN. As a result, every UE has exactly one tunnel per PDN where only one QoS can be supported. Beside these QoS limitations with PMIPv6 the UEs always have to be on, since there is no flow control with PMIPv6. This results in continuous energy consumption at the UE. The PMIPv6 standard and extensions of the standard do not define any solution to put the UE into idle or sleep mode and put the UE back into active mode if there is some data ready to send to the UE. With GTP the UE is able to go into idle mode and go into active mode if there is data destined for the UE because of the supported paging mechanism.

4.4.1 GTP and PMIPv6 in the Core Network

The two network-based mobility protocols, the GTP and the PMIPv6, can be used in the core network on the S5/S8 interface. The choice of the mobility protocol on the S5/S8 interface has impacts on the architecture as well as on the QoS mechanisms and the amount of established tunnels. The specifications for the GTP-based S5/S8 interface are given in (3GPP TS 23.401 V12.3.0, 2013), whereas the PMIPv6-based variant is specified in (3GPP TS 23.402 V12.3.0, 2013).

Within the EPC, both mobility protocols can be used. Figure 26 shows which mobility protocol can be deployed in which area.

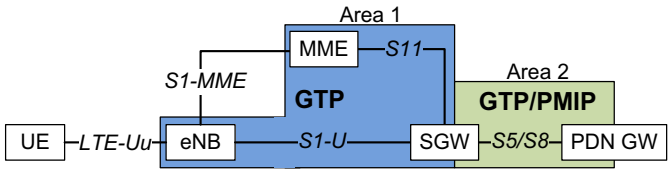


Figure 26: EPC with deployed mobility protocols

On the S1-U and the S11 interface GTP is the only allowed mobility protocol (see Figure 26, area 1). On the S5/S8 interface (see Figure 26, area 2) either GTP or PMIPv6 can be deployed. Regardless of the protocol used on the S5/S8 interface, GTP is always used towards the eNB (see Figure 26 area 1). This fact eliminates some of

the functional drawbacks of the PMIPv6 which are identified in Table 7. Which functional drawbacks of the PMIPv6 can be compensated by the mandatory use of GTP in area 1 is shown in Table 8.

Table 8: Functions supported by the areas with deployment variations of the mobility protocol

	Area 1=GTP	Area 2=GTP	Area 2=PMIPv6
QoS signalling	+	+	-
Bearer signalling	+	+	-
Packet forwarding during handover	+	Is performed in area 1	Is performed in area 1
Mobility tracking	+	Is performed in area 1	Is performed in area 1
Paging	+	Is performed in area 1	Is performed in area 1

Since area 1 provides GTP functionality, the performance of the handover process can be improved as well as the power consumption behaviour of the UE, through mobility tracking and paging functions, provided by area 1. But the QoS and bearer signalling functionalities are still not available on the PMIPv6-based S5/S8 interface.

The architecture and the functionality of the EPC entities vary, subject to the deployed mobility protocol on the S5/S8 interface. The consequences which result from the different mobility protocols are outlined in the following. The architecture of the EPC including the interworking towards the PCC when GTP is deployed on the S5/S8 interface, is shown in Figure 27. At the PDN GW, there are several SDFs for transfer in downlink direction. Through downlink Traffic Flow Templates (TFTs), the SDFs are mapped on the appropriate EPS bearer, which supports a specific QoS and is represented by one GTP tunnel on the interfaces S5/S8, S1-U, as well as by the LTE-Uu interface. One GTP tunnel represents a bearer together with a QoS class.

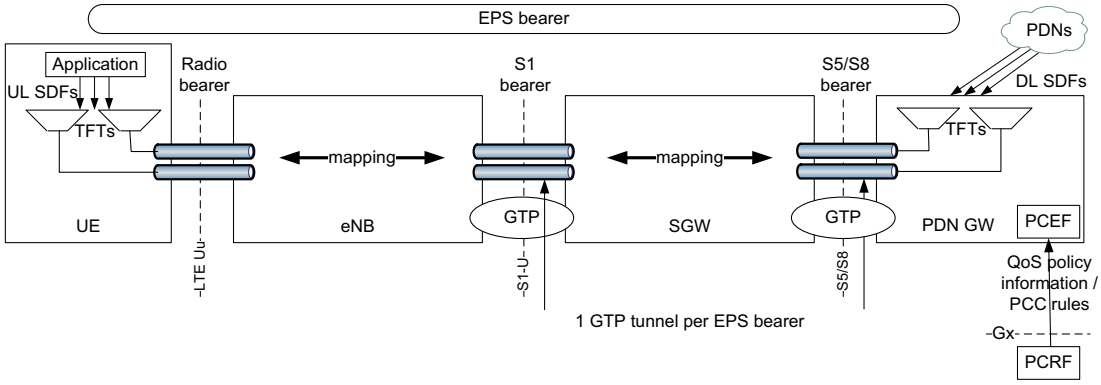


Figure 27: Architectures for GTP-based S5/S8 interface

The filters used for the mapping are [source_IP, destination_IP, source_port, destination_port, transport_protocol_ID] IP-5-tuples. The TFTs are provided by the PCRF towards the PCEF within the QoS policy information, the PCC rules. The PCEF is

located in the PDN GW. Within the SGW and the eNB a 1 to 1 mapping between the different bearers in both up- and downlink direction is applied. The UE receives the same QoS policy information, the PCC rules, like the PCEF. The QoS policy information sent to the UE within the GTP includes the uplink TFTs to map the traffic to be sent towards the appropriate radio bearer. The EPS bearer with the appropriate QoS spans from the UE to the PDN GW.

When PMIPv6 is deployed on the S5/S8 interface, the architecture changes significantly compared with the GTP-based S5/S8 interface. A major difference is the shortened EPS bearer. It reaches only from the UE via the eNB to the SGW as can be seen in Figure 28. The S5/S8 bearer between the SGW and the PDN GW is no longer part of the EPS bearer. There are filters required at both EPC entities, the PDN GW and the SGW.

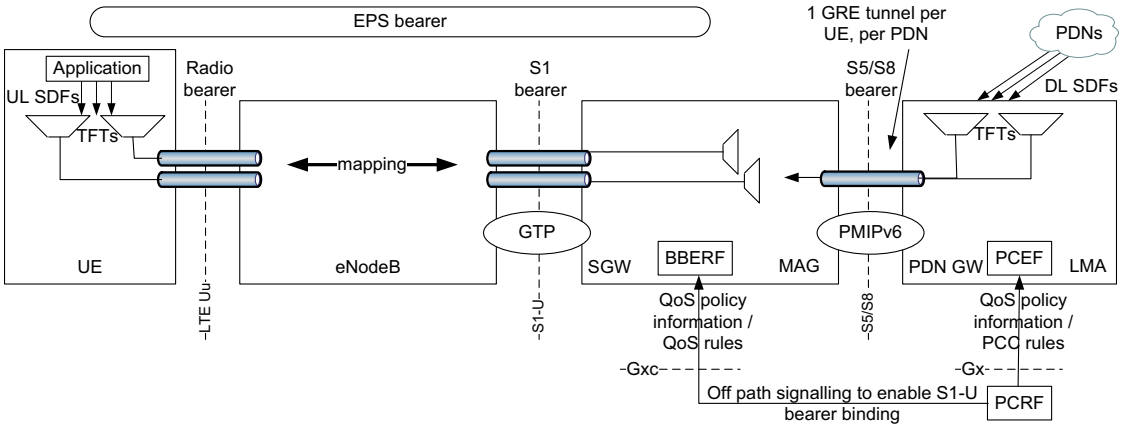


Figure 28: Architectures for PMIPv6-based S5/S8 interface

The SGW takes over the bearer binding functionality with PMIPv6 on the S5/S8 interface. With GTP the bearer binding is performed by the PDN GW. As a result of this new functionality in the SGW, the PCRF must provide the necessary QoS policy information, the QoS rules, towards the Bearer Binding and Event Reporting Function (BBERF) entity. The BBERF is similar to the PCEF but located within the SGW and only used, if PMIPv6 is applied on the S5/S8 interface. The QoS policy information is necessary to enable the SGW to map the incoming SDFs towards the appropriate EPS bearer. The reasons to still apply filters in the PDN GW are on one hand because the charging is managed in the PDN GW and on the other hand because traffic shaping is performed by the PDN GW. But the filters on the PDN GW have no relevance on the bearer binding anymore. To operate PMIPv6 on the S5/S8 interface the Local

Mobility Anchor (LMA) is located at the PDN GW and the Mobility Access Gateway (MAG) is located at the SGW. Between the PDN GW and the SGW GRE tunnels are established to separate the traffic for each UE and PDN. With this separation it is possible for a UE to perform a handover per PDN. Assumed that a UE has connections to multiple PDNs (e.g. Internet and IMS), it is possible to independently do a handover for a single PDN connection per UE. The basic standard of PMIPv6 does not provide support of IP flow mobility. To add this capability, the IETF defined in (IETF, NETEXT Working Group, 2013) the ability that a mobile node can connect to the same PMIPv6 domain through different interfaces and in (IETF, NETEXT Working Group, 2013) the ability that a MAG can forward traffic to a mobile device, even if the IP address was originally delegated to the mobile node via a different MAG. The GRE tunnels do not support any QoS signalling.

All the necessary QoS interactions from the gateways with the PCC are defined for both mobility protocols used on the S5/S8 interface. The signalling effort from GTP and PMIPv6 during a handover has been evaluated. This evaluation of the signalling effort for both GTP and PMIPv6 mobility protocols is done to select the preferred mobility protocol on the S5/S8 interface. The choice of the mobility protocol in area 2 on the S5/S8 interface does not have any impact on the signalling effort in area 1. Therefore, the analysis of the signalling effort is done only for the area 2 where either of the two mobility protocols can be deployed. Two scenarios are defined:

Scenario 1 (sc1): GTP is deployed on the S5/S8 interface

Scenario 2 (sc2): PMIPv6 is deployed on the S5/S8 interface.

In the following the signalling effort for an S1 handover with SGW relocation in both scenarios is provided. The S1 handover with SGW relocation situation is shown in Figure 29.

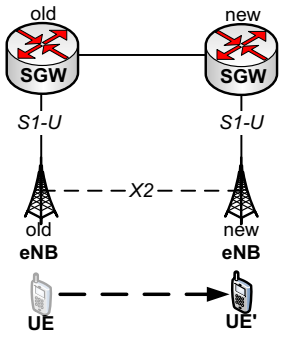


Figure 29: S1-based handover scenario with SGW relocation

The signalling load includes the necessary exchange of signalling information above the layer 2. IPv6 is used as the layer 3 protocol. The following equations use the abbreviation *tse* for the *total signalling effort* above layer 2. The unit of *tse* is Byte. Furthermore a signalling effort analysis is provided for both scenarios where multiple bearers are used. For the comparison only the successful cases are considered.

The signalling effort on the S5/S8 interface for an S1 handover, using GTP as the mobility protocol, contains the GTP Modify Bearer Request (MBReq) which is sent from the SGW to the PDN GW and the GTP Modify Bearer Response (MBResp) sent from the PDN GW to the SGW. For calculating the signalling effort, the sequence diagrams of the 3GPP standard (3GPP TS 23.401 V12.3.0, 2013) were used to identify the relevant messages and their Information Elements (IEs). Furthermore, the 3GPP standard (3GPP TS 29.274 V12.3.0, 2013) was used to calculate the message sizes on the basis of the identified IEs. Since the signalling effort is defined as the whole signalling effort above layer 2 the User Datagram Protocol (UDP) (IETF RFC 768, 1980) and IPv6 (IETF RFC 2460, 1998) headers (UDP_hdr, IP_hdr) have to be added to the signalling effort of the GTP message, resulting in equation (3) for the signalling effort of the MBReq message.

$$tseMBReq=MBReq+IP\_hdr+UDP\_hdr=62+40+8=110 \quad (3)$$

The MBResp message has the signalling effort according to equation (4).

$$tseMBResp=MBResp+IP\_hdr+UDP\_hdr=44+40+8=92 \quad (4)$$

The whole signalling effort for both messages is 202 Bytes.

The procedures, which have an impact on the signalling effort for an S1 handover when PMIPv6 is used on the S5/S8 interface, are shown in Figure 30. First the SGW has to establish a control session towards the PCRF over the Gxc interface to get the QoS policy information required for the bearer binding at the SGW.

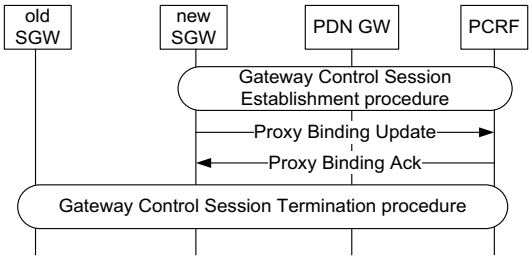


Figure 30: Handover procedure with PMIPv6-based S5/S8 interface

Then the proxy binding update with the appropriate response can be processed and finally the old SGW terminates its session towards the PCRF. To get the required QoS information, the SGW performs a Gateway Control Session Establishment (GCSE) procedure towards the PCRF. The SGW sends a Credit Control Request (CCR) Diameter message (CCR_msg) to the PCRF and receives a Credit Control Answer (CCA) Diameter message (CCA_msg) back from the PCRF, which contains all the necessary QoS information to enable the SGW performing the bearer binding. The layer 4 protocol for the CCR_msg and the CCA_msg is the Stream Control Transmission Protocol (SCTP). Therefore, the header and chunk (SCTP_hdr_chunk) of the SCTP have to be added to get the total signalling effort for the CCR_msg and CCA_msg. The calculations of the signalling efforts have been made based on the sequence diagrams and the message descriptions of the 3GPP standards (3GPP TS 29.213 V12.2.0, 2013), (3GPP TS 29.212 V12.3.0, 2013), (3GPP TS 23.402 V12.3.0, 2013), (3GPP TS 23.203 V12.3.0, 2013), the RFC definitions of the diameter protocol (IETF RFC 3588, 2003), (IETF RFC 4006, 2005), the PMIPv6 definitions (IETF RFC 5213, 2008), (IETF RFC 3775, 2004), (IETF RFC 5149, 2008), (IETF RFC 5845, 2010), (3GPP TS 23.003 V12.1.0, 2013), the IPv6 definition (IETF RFC 2460, 1998), and the SCTP definition (IETF RFC 4960, 2007). The signalling effort for requesting the QoS information from the PCRF is given by equation (5).

$$tseGCSEReq=CCR\_msg+SCTP\_hdr\_chunk+IP\_hdr=336+28+40=404 \quad (5)$$

The signalling effort to provide QoS information from the PCRF to the SGW is 1040 Bytes according to equation (6).

$$tseGCSEResp=CCA\_msg+SCTP\_hdr\_chunk+IP\_hdr=972+28+40=1040 \quad (6)$$

The tse of the whole GCSE procedure results in 1444 Bytes. Following the GCSE procedure, a Proxy Binding Update (PBU) message will be sent from SGW to the PDN GW according to equation (7), and then the Proxy Binding Acknowledgement (PBA) message will be sent from the PDN GW to the SGW according to the equation (8).

$$tsePBU=PBU+IP\_hdr=104+40=144 \quad (7)$$

$$tsePBA=PBA+IP\_hdr=104+40=144 \quad (8)$$

Finally, to terminate the session between the old SGW and the PCRF a Gateway Control Session Termination (GCST) procedure is performed. The termination of the session is initiated by the SGW sending a CCR Diameter message to the PCRF and the PCRF responds with a CCA Diameter message. The identified signalling efforts for the session termination request and the response are shown in equation (9)and (10).

$$tseGCSTReq=CCR\_msg+SCTP\_hdr\_chunk+IP\_hdr=212+28+40=280 \quad (9)$$

$$tseGCSTResp=CCA\_msg+SCTP\_hdr\_chunk+IP\_hdr=188+28+40=256 \quad (10)$$

The percentage of the signalling effort per procedure for an S1 handover with PMIPv6 is shown in Figure 31. The PBU and the PBA procedures have the lowest impact, followed by the GCST procedure. The highest impact is caused by the GCSE procedure, which provides the QoS information.

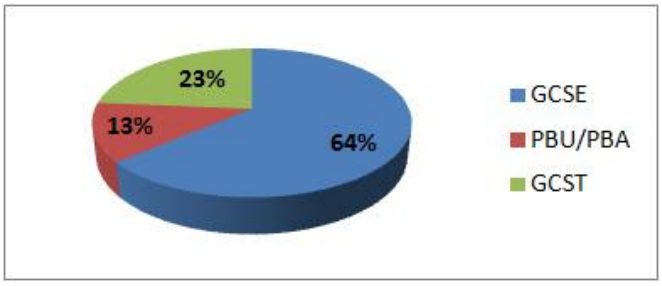


Figure 31: Percentage of the signalling effort per procedure for an S1 handover with PMIPv6

The overall signalling effort for the required procedures, with PMIPv6 as the mobility protocol, is 2268 Bytes. This is a factor 11.2 increase of the signalling effort for PMIPv6 compared with the signalling effort that GTP requires. The QoS information has the highest impact on the overall signalling effort when PMIPv6 is used. This impact increases even more, if multiple bearers are used, as is shown in the following.

A UE can hold a maximum of 11 bearers. This value affects both scenarios, but the GTP and PMIPv6-based S5/S8 interfaces show different additional signalling effort (*se*) for an additional bearer. In sc1 the additional information for one bearer consists of the bearer context according to equation (11), which is an IE within the GTPv2-C.

$$\Delta s1\_se\_per\_bearer=bearer\ context=43\ Bytes \quad (11)$$

The sc2 requires an additional QoS-Rule-Install Attribute Value Pair (AVP) within the CCA diameter message for one additional bearer as according to equation (12).

$$\Delta s2\_se\_per\_bearer = QoS\text{-}Rule\text{-}Install\ AVP = 792\ Bytes \quad (12)$$

If PMIPv6 is the used mobility protocol the additional signalling effort for an extra bearer is 18.4 times higher than with GTP. A number of 4 to 6 bearers will be common. For example, assuming a UE has 2 connections to different PDNs. One connection is towards the IMS, the other connection is towards the internet. One default bearer is needed for each PDN. Furthermore it is assumed that the UE has 2 dedicated bearers for the IMS, one for voice service and the other one for a video streaming service, as well as one dedicated bearer for the internet PDN. For this example the UE holds a number of 5 bearers. Equation (13) is used to calculate the total number of bearers:

$$total\_bearer\_number = PDN\_number + dedicated\_bearer\_number \quad (13)$$

An overview of the establishment procedures of default and dedicated bearer including a signalling effort evaluation is given in (Frei et al., 2009). Figure 32 shows the percentage of the signalling effort of the GCSE procedure within the sc2 in relation to the overall signalling effort of the sc2 for an increasing number of bearers. The impact of the GCSE procedure, where the session from the SGW towards the PCRF is established and the needed QoS information is sent to the SGW increases from 64 % (one bearer) to 92 % (11 bearers).

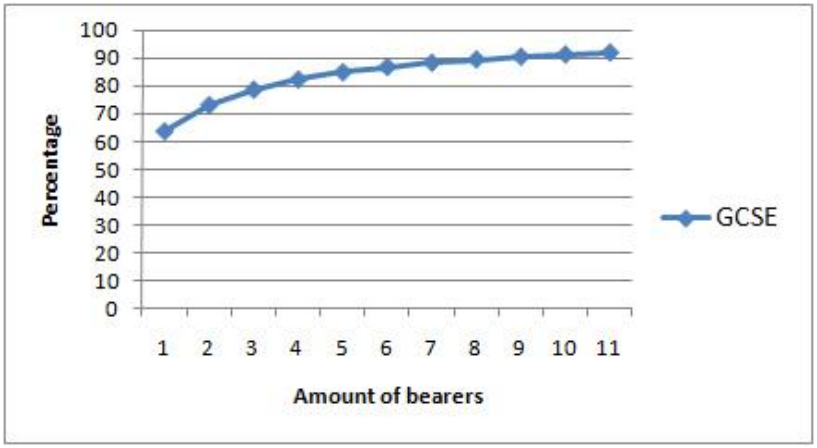


Figure 32: Percentage signalling effort of GCSE to overall signalling effort

The impact of the QoS information on the signalling effort is significant. And this impact occurs not only during handover, but also during bearer modification, QoS modification and if a dedicated bearer is activated.

For the QoS enforcement on layer 3 the differentiated service (DiffServ) (IETF RFC 2474, 1998) mechanism is chosen. DiffServ uses classes to differentiate the services, also called Class of Service (CoS). To divide the packets into different classes, the Type of Service (ToS) byte in the IP version 4 (IPv4) and the Traffic Class field in the IPv6 are used. It is possible to statically map the QCI values to Differentiated Service Code Points (DSCPs). Since the QCI values are dynamically assigned to the bearers, the provided QoS on IP layer will still be dynamic, even if the mapping from QCI to DSCP values is done in a static way.

Figure 33 shows the QoS enforcement of the EPS QoS with DiffServ if GTP is deployed on the S5/S8 interface. The uplink direction is shown in green and the downlink direction is shown in blue.

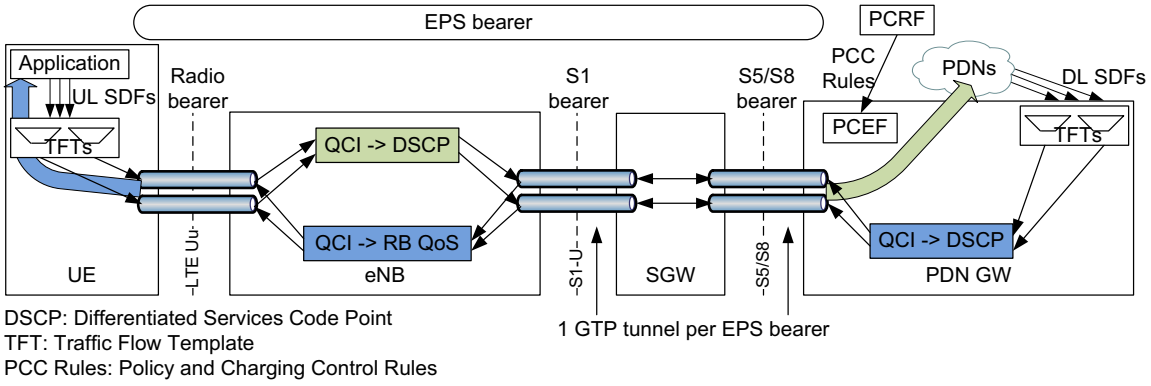


Figure 33: QoS enforcement with GTP based S5/S8 interface

The pending packets of the downlink SDFs are mapped to the appropriate bearers by applying the TFTs at the PDN GW. Then the assigned QCI values are mapped statically to the DSCP values and the packets are sent through the appropriate GTP tunnel. Within the SGW there is a 1 to 1 mapping of the S5/S8 GTP tunnel and the S1 GTP tunnel in both directions (uplink and downlink). The eNB applies a mapping of the QCI values towards the Radio Bearer (RB) QoS values in downlink direction. Finally, the UE forwards the packets to upper layers until the packets reach the application. The uplink SDFs, generated by applications, are mapped using the TFTs in the UE to the appropriate RB. The eNB performs a static QCI to DSCP mapping on the uplink

traffic. There is also a 1 to 1 mapping in the SGW for the uplink direction. The PDN GW routes the packets to the destined PDNs.

Figure 34 shows the QoS enforcement of the EPS QoS and DiffServ QoS, if PMIPv6 is deployed on the S5/S8 interface. As in Figure 33 the uplink direction is shown in green and the downlink direction is shown in blue within Figure 34. The bearer binding is performed at the SGW in the BBERF, since PMIPv6 is deployed on the S5/S8 interface. Therefore, the PCRF has to provide the PCC Rules to the PCEF (to enable charging and gating functions) and the QoS rules to the BBERF. The additional signalling effort, to distribute the QoS rules, has been evaluated in (Frei et al., 2010) and the result was 1444 Bytes. This additional signalling effort appears every time a bearer is modified, activated or a handover is performed.

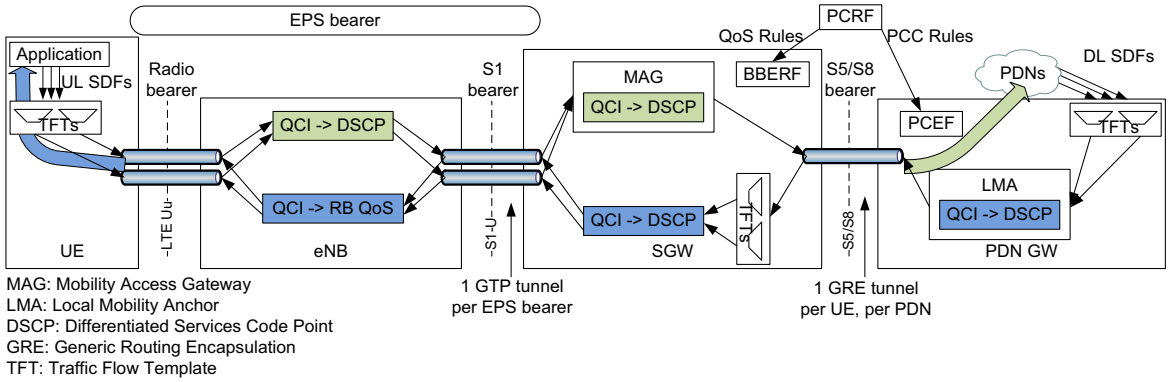


Figure 34: QoS enforcement with PMIPv6 based S5/S8 interface

One significant difference between the two mobility protocol architectures is that there is no EPS bearer present at the S5/S8 interface and therefore there is not one QoS level per tunnel. Instead of one QoS level per GTP tunnel there is one GRE tunnel per UE and per PDN only. Therefore, one GRE tunnel is assigned to multiple QoS levels. As a result of this, the QCI to DSCP mapping is much more complex with PMIPv6. Because the QCI values and the TFTs are also known in the PDN GW, even if the PMIPv6 is deployed on the S5/S8 interface, the operator defined QCI to DSCP mapping is also possible in this deployment scenario through DSCP marking before the tunnelling encapsulation takes place. The process of DSCP marking before the encapsulation takes place is called QoS pre-classification. This is done by copying the packet before encryption takes place, and afterwards adding the QoS values from the copied packet to the encrypted one. This can be done in the LMA for downlink and in the MAG for the uplink traffic. In the downlink direction, the bearer binding is

performed in the SGW through the TFTs and an additional QCI to DSCP mapping is processed. The eNB and the UE behaviour are the same with PMIPv6 than with GTP as the macro mobility protocol.

This analysis has shown that it is possible to enforce DiffServ on the IP layer with both macro mobility protocols. If PMIPv6 is used, an additional signalling effort occurs due to the additional distribution of the QoS rules from the PCRF to the BBERF and the effort of performing the QCI to DSCP mapping is higher with PMIPv6 than with GTP, because with PMIPv6 the QCI to DSCP mapping has to be performed 4 times compared to 2 times when GTP is used. Furthermore, it has been shown that the deployment of the PMIPv6 protocol causes more signalling effort than the deployment of GTP on the S5/S8 interfaces. The QoS information is required to be sent from PCRF to the SGW during: S1 handover with SGW relocation, bearer modification, QoS modification, and if a dedicated bearer is activated. The conclusion of these signalling effort evaluations is that GTP performs clearly better than PMIPv6 during the S1 handover and whenever QoS information has to be sent from the PCRF to the SGW. The signalling effort for a handover with the PMIPv6 protocol is much higher with PMIPv6 than with GTP. Furthermore, with GTP the QoS enforcement requires less effort at the network nodes than with PMIPv6. These results are also confirmed through the fact that the GTP protocol has been selected beside the PMIPv6 as the alternative mobility protocol for non-3GPP access on the S2b interface since the 3GPP Release 10 defined in the technical report (3GPP TR 23.834 V10.0.0, 2010) and also on the S2a interface since the 3GPP Release 11 defined in the technical report (3GPP TR 23.852 V12.0.0, 2013). The S2b interface is used for the integration of the non-trusted, non-3GPP access networks whereas the S2a interface is used for the integration of trusted non-3GPP access networks. In former releases only PMIPv6 was allowed on the S2a and S2b interfaces. This development indicates that the GTP protocol gains more influence than the PMIPv6 protocol.

These research results and the more extensive support for the GTP protocol through the 3GPP standardisation has led to the decision that GTP is selected as the recommended mobility protocol for the S5/S8 interface throughout this thesis.

4.4.2 GTP and PMIPv6 Between the Access and the Core Network

GTP and PMIPv6 can both be used on the S2a and S2b interfaces to enable non-3GPP trusted and non-trusted access networks the access towards the EPC. Figure 35 shows how the non-3GPP access networks are gaining access towards the EPC.

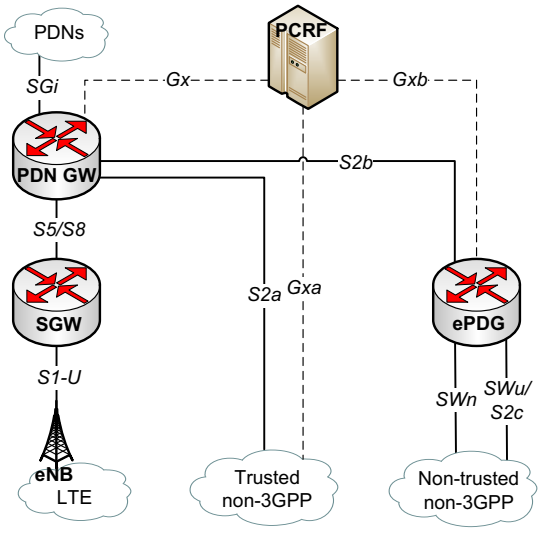


Figure 35: Non-3GPP access with GTP and PMIPv6 towards the EPC

Similar to the situation in the core network, there is also an additional signalling effort with PMIPv6-based S2a and S2b interfaces to provide the policies from the PCRF to the ePDG through Gxb interface and to the trusted WLAN through Gxa interface. This additional signalling effort is not required when using GTP. However, the Gxa interface is not used in 3GPP Release 12 (3GPP TS 23.402 V12.3.0, 2013). The inability to use the Gxa interface has a significant impact on the granularity of the QoS enforcement. To show this impact both variants, with the Gxa interface and without it, are discussed in the following. Even if the Gxb interface is used in the 3GPP standards and no restrictions are applied to the Gxb interface the impact of not using the Gxb interface is discussed as well in the following. The solutions using the Gxa and the Gxb interface are shown in light grey, and similar to the figures in section 4.4.1, the downlink direction is shown in blue whereas the uplink direction is shown in green in the following figures.

Figure 36 shows how the QoS is enforced with a GTP-based S2b interface. At the PDN GW the TFTs are applied to the downlink SDFs for bearer binding and the traffic is classified by different QCIs. Afterwards, the QCI to DSCP mapping is applied and the data packets are sent into the appropriate GTP tunnel, which represents an EPS

bearer. At the ePDG the data packets are extracted by eliminating the outer IP header and the GTP header. Afterwards, the QCI to DSCP mapping has to be done either one more time, because the DSCP information added at the PDN GW in the outer IP header has been removed or the DSCP value from the outer IP header has to be copied and inserted into the IP header. In the uplink direction the TFTs are applied at the ePDG to realise the bearer binding and afterwards the QCI to DSCP mapping is performed analogous to the procedure at the PDN GW in downlink direction.

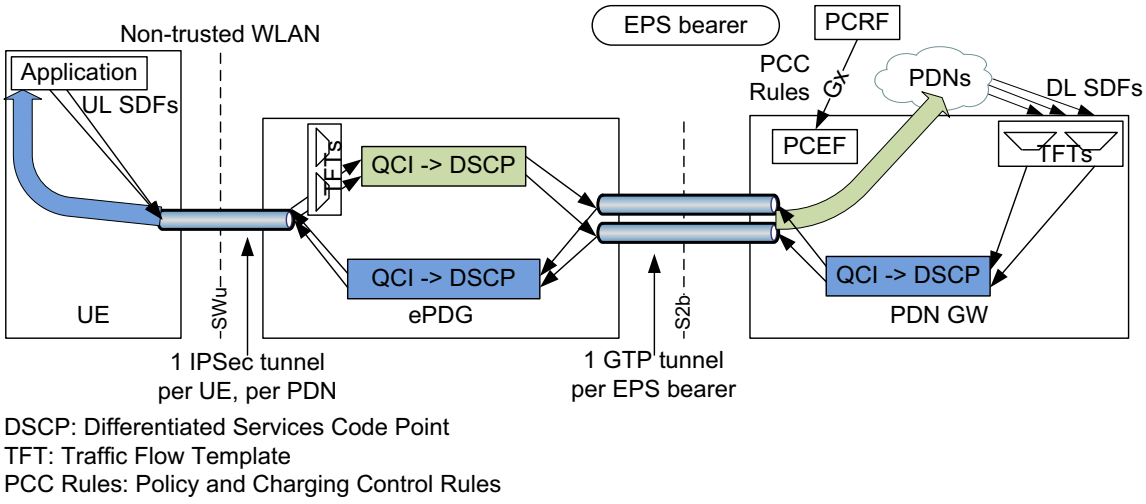


Figure 36: DiffServ with GTP-based S2b interface

Figure 37 shows the DiffServ QoS enforcement with a PMIPv6-based S2b interface. The QoS enforcement is the same with PMIPv6 at the PDN GW as with GTP. If the Gxb interface is used the QoS enforcement will be the same with PMIPv6 for the uplink data traffic at the ePDG as with GTP. But it is slightly different for the downlink data traffic at the ePDG. The TFTs have to be applied before the QCI to DSCP mapping is preceded because the DSCP information from the outer header from the GRE tunnel is not available anymore and furthermore, all traffic per PDN and per UE is sent over the same GRE tunnel and therefore the traffic cannot be differentiated with the tunnel information, because there is only one tunnel. As a result the TFTs have to be applied also at the ePDG to enable the QCI to DSCP mapping. The QoS enforcement is totally different if the Gxb interface is not used. As a result of the inability to use the Gxb interface the policies cannot be provided from the PCRF towards the ePDG with a PMIPv6-based S2b interface. This leads to the situation that there are no TFTs present at the ePDG to classify the traffic into different QCIs. As a result, the DSCP marking has to be applied at the ePDG for both uplink and downlink data traffic. To

realise a DSCP marking, filters are also required to classify the data traffic. The commonly applied filters for DSCP marking are nowadays static ones. As a result of static configured filters, the DSCP marking is as well static and therefore the QoS enforcement is courser-grained compared to the fine-grained and dynamic QoS enforcement which is possible with the use of the Gxb interface or with a GTP-based S2b interface.

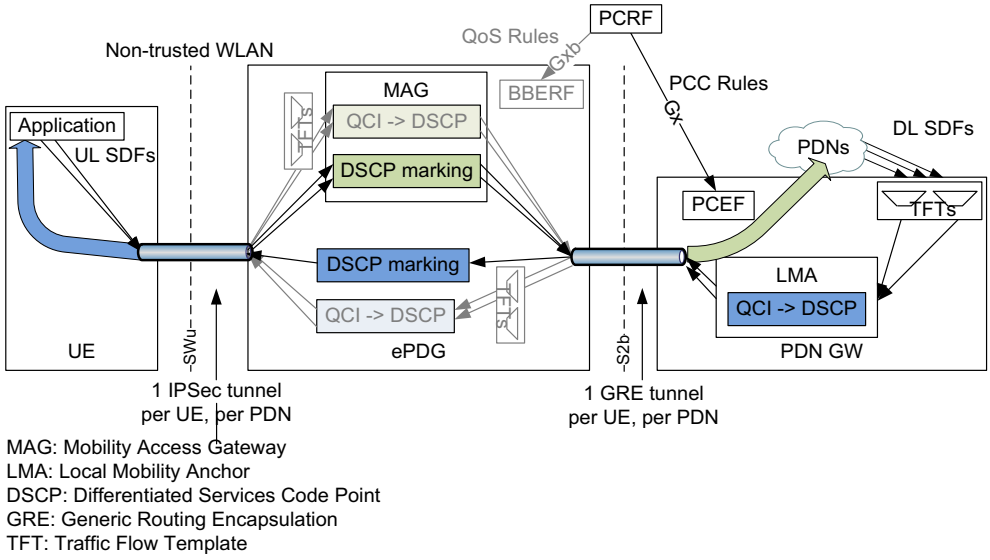


Figure 37: DiffServ with PMIPv6-based S2b interface

The QoS enforcement using DiffServ is also different on the S2a interface depending on the used mobility protocol. Figure 38 shows the QoS enforcement with DiffServ with a GTP-based S2a interface. At the PDN GW the DiffServ enforcement with the GTP-based S2a interface is the same as with the GTP-based S2b interface. First, the TFTs are applied and afterwards the QCI to DSCP mapping is processed. At the trusted WLAN access network the QCI to DSCP mapping is processed for the downlink data traffic. The uplink data traffic at the trusted WLAN is first classified into different QCIs applying first the TFTs to perform the bearer binding and then the QCI to DSCP mapping. The QoS mapping from the DSCP values to the radio QoS values from the 802.11 standards and vice versa at the AP is not shown in Figure 38 and Figure 39.

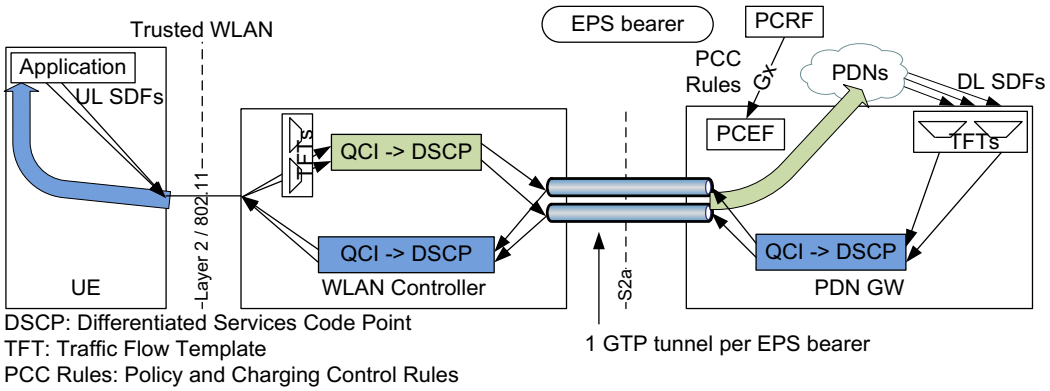


Figure 38: DiffServ with GTP-based S2a interface

Figure 39 shows how the QoS is enforced with DiffServ and a PMIPv6-based S2a interface with and without the use of the Gxa interface. Both solutions, with and without the Gxa interface are in principle the same for both, uplink and downlink direction as it is the case with the PMIPv6-based S2b variant with the difference that the DiffServ enforcement is not located at the ePDG, instead it is located at the WLAN controller. The solution without the Gxa interface results again in a static DSCP marking and therefore the QoS enforcement is courser-grained compared to the fine-grained dynamic QoS enforcement which is possible with the use of the Gxa interface or with a GTP-based S2a interface.

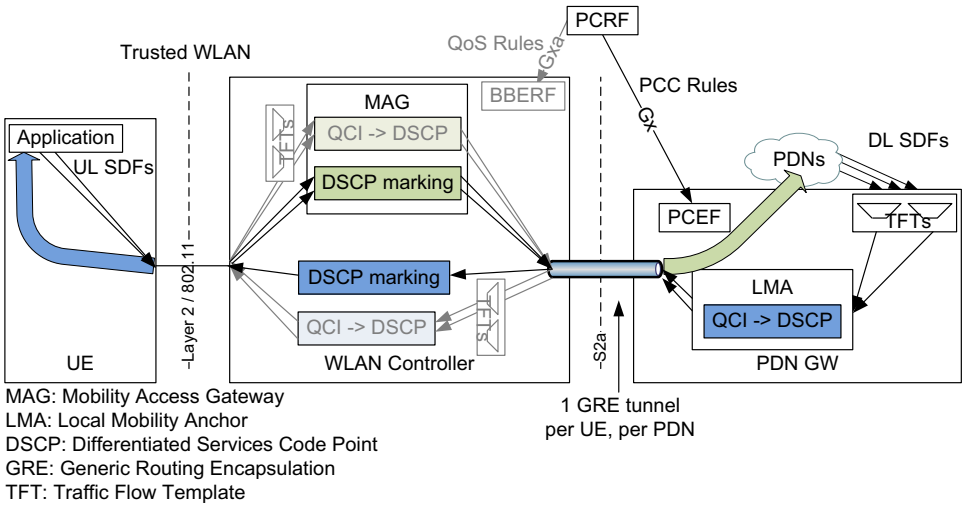


Figure 39: DiffServ with PMIPv6-based S2a interface

The QoS can be enforced generally with less effort with GTP than with PMIP. The GTP-based S2a and S2b interfaces perform much better than PMIP-based S2a and S2b interfaces without the Gxa and the Gxb interfaces, because the QoS enforcement can be done in a dynamic way which results in a finer-grained QoS. Even with the use of

the Gxa and Gxb interfaces the GTP-based S2a and S2b can enforce the DiffServ in an easier way than it is possible with PMIPv6-based S2a and S2b interfaces. Both network-based mobility protocols support the offloading of traffic through MAPCON and IFOM. GTP allows the UE to switch into idle mode to save energy at the UE. This is possible because GTP supports paging. The idle mode is not defined in the PMIPv6 standard, because there is no possibility of paging the UE. Because of these advantages that can be exploited with GTP as the mobility protocol GTP is also the selected network-based mobility protocol between the access and the core network.

4.5 Interface to Enable the Access to the EPC for WLANs

WLAN can get access to the 3GPP´s EPC over three interfaces: the S2a, the S2b, and the S2c interface. Table 9 gives again an overview whether the S2x interfaces support trusted and/or non-trusted non-3GPP access and which mobility protocol is used. Since the DSMIPv6 is a host-based mobility protocol and in this thesis a network-based mobility protocol has been selected, the S2c interface was dropped and only the S2a and S2b interfaces are considered in this thesis. The choice for the network-based mobility protocol is GTP and therefore PMIPv6 is not considered further. The green highlighted fields in the table indicate the further considered interfaces and protocols in this thesis.

Table 9: Trusted and untrusted access of the S2x interfaces with the corresponding mobility protocols

	S2a	S2b	S2c
Mobility protocol	PMIPv6	PMIPv6	DSMIPv6
	GTP	GTP	
Trusted	X		X
Non-trusted		X	X

Therefore, the decision is only, which of the interfaces S2a or S2b is selected in this thesis. Figure 40 shows the two possibilities of providing access to the EPC for trusted and non-trusted WLAN. The non-trusted WLAN gets access to the PDN GW through the ePDG. Between the UE and the ePDG, an IPsec tunnel is established. The drawback of this solution is that there is only one IPsec tunnel per UE and per PDN. As a result, only one QoS class can be enforced per PDN. To enable fine-grained QoS marking both the UE and the ePDG would have to support QoS pre-classify mechanisms which enable to copy the original packet before encryption takes place, and afterwards adding the QoS values from the copied packet to the encrypted one. Between the ePDG and the PDN GW there are multiple GTP tunnels established. Each GTP tunnel

supports a different QoS class to realise an S2b bearer. Therefore, the QoS enforcement is fine-grained on the S2b interface and coarse-grained on the SWu interface, except for the case that the UE and the ePDG both support QoS pre-classification. On the other hand, the trusted WLAN can obtain direct access to the PDN GW without an additional gateway in between. The WLAN QoS mechanisms (IEEE 802.11e, 2005) can be used between the UE and the WLAN access point. Between the WLAN and the PDN GW 3GPPs fine-grained QoS concept can be used on the S2a interface with one GTP tunnel per bearer.

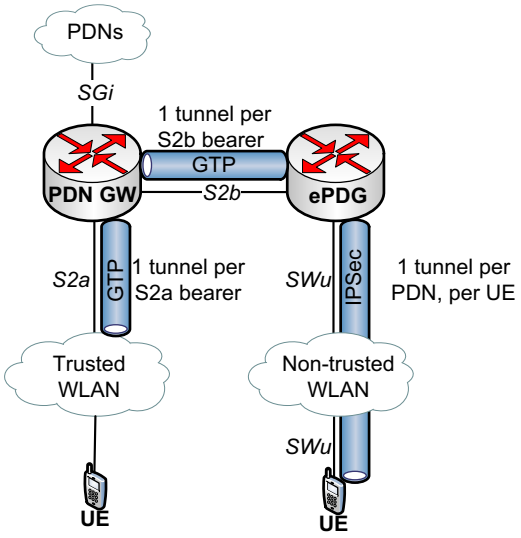


Figure 40: Provision of EPC access to WLAN

The EPC interworking with a trusted WLAN is generally an easier procedure than with a non-trusted WLAN, since the IPsec tunnel produces additional overhead. Furthermore, the QoS can be enforced easier with a trusted WLAN and a fine-grained QoS can be enforced at all times. Therefore, the S2a interface to enable the access for trusted WLAN access is selected for further considerations in this thesis.

4.6 The Need for Policies

The predicted huge increase of traffic load (Cisco, 2014) requires new functionalities and cannot be handled only with LTE network upgrades. Therefore, appropriate steps towards the solving or reduction of the impact of this problem have to be taken. Several proposed capability enhancements towards solving the capacity problem are given in the following:

- Offload data traffic to available HetNets.

- Efficient routing in the EPC network, for example SIPTO.
- Steering data traffic within the available network resources.
- React dynamically on changing load levels.
- Enforce different levels of user subscription.
- Enforce different QoS classes.
- Gating control with the ability of blocking specific traffic.
- Appropriate network discovery
- Appropriate network selection mechanisms

As can be seen from the list above, the problems that have to be solved are complex and sometimes even the different capabilities compete with one another. A key enabler to realise these capabilities and solve conflicts between them is the use of policies. The three central policy nodes in the PCC are the PCRF, which maintains, supervises, and creates policies per user, the PCEF (GTP on the S5/S8 interface), and the BBERF (PMIPv6 on the S5/S8 interface), respectively, which enforces the policies. The PCRF, and the PCEF/BBERF have a restrictive approach with policies. As a result, the policies are not only recommendations, they are rules that are strictly applied to user data traffic and therefore these policies are binding. The ANDSF is another policy related node which provides policies towards UEs, but in contrast to the PCC policies, the ANDSF policies are recommendations that assist the UE for example in network discovery, with the selection of the access network, traffic offloading etc. The Wi-Fi Alliance specified in its Hotspot 2.0 Release 1 improvements for end-user devices to use nearby Wi-Fi access points, as well as to discover and authenticate to access points in an automatic way and without an intervention from the end-user. At the time of writing this thesis, the Hotspot 2.0 Release 2 is being drafted. The main changes would be the addition of operator controlled policies. All these solutions, the PCC, the ANDSF, and the Hotspot 2.0 Release 2 make use of policies to tackle the capacity problem. This thesis proposes further new and advanced solutions making also use of policies.

4.7 IEEE´s MIHS or 3GPP´s ANDSF Solution

Both, the IEEE MIHS (IEEE 802.21, 2009) and the 3GPP ANDSF (3GPP TS 24.302 V12.3.0, 2013), (3GPP TS 23.402 V12.3.0, 2013) standards specify solutions to enable

inter-system mobility in HetNets. Table 10 compares the 3GPP solution with the ANDSF in combination with the EPC functionalities with the IEEE solution the MIHS.

Table 10: Similarities and contrasts of the IEEE and the 3GPP framework

<table>
<tr><th colspan="2">IEEE MIHS</th><th>3GPP ANDSF/EPC</th></tr>
<tr><td colspan="2">MIES
Events are sent between UE and network node.</td><td>The Event Reporting Function of the EPC network is comparable to the MIES. It is located either in the PCEF or in the BBERF (depends on the deployed mobility protocol) and report events to the PCRF. Both, the PCEF and the PCRF are network nodes.</td></tr>
<tr><td colspan="2">MICS</td><td>The EPC has also a mechanism to reserve resources but the MICS provide a wider range of commands than the EPC.</td></tr>
<tr><td>MIIS</td><td>The information services are similar to each other.</td><td>Access network discovery information</td></tr>
<tr><td colspan="2">A mechanism similar to the inter-system mobility policy is not supported within the IEEE MIH.</td><td>Inter-system mobility policy</td></tr>
<tr><td colspan="2">A mechanism similar to the inter-system routing policy is not supported within the IEEE MIH.</td><td>Inter-system routing policy</td></tr>
</table>

All the capabilities provided by the MIHS are supported by the ANDSF in combination with the EPC. The MICS provides a wider range of commands than the EPC. But the ANDSF provides two kinds of policies, the inter-system mobility policy to provide operator defined policies and also restrictions and the inter-system routing policy to enable simultaneous routing over multiple interfaces. Both kinds of policies are not supported with the MIHS.

The paper (Frei et al., 2011b) discusses the possibilities of combining the MIHS with the ANDSF and EPC functionalities to eliminate several drawbacks. 3GPP decided in Release 8 that the MIHS will not be further considered for inter-system handover in the EPS. Instead they defined their own network element, the ANDSF to support inter-system handover as well as traffic offloading. With the definition of the Wi-Fi Alliance Hotspot 2.0 specification there is a comprehensive solution to enable the collaboration of Wi-Fi and the EPS. In the standardisation work of 3GPP the Hotspot 2.0 specification is already integrated and it is expected that co-operation of 3GPP and Wi-Fi Alliance will continue in the future. Therefore, the IEEE MIHS solution is not further considered within this thesis.

4.8 Traffic Steering Control and Decision Point

It is reasonable to realise the traffic steering control rather from the network side than from the terminal side. The reasons selecting the network-controlled terminal assisted approach for this thesis are listed in the following:

- The 3GPP cellular networks use a network-controlled terminal assisted approach for the defined basic traffic steering mechanisms such as handover.
- Traffic steering needs algorithms to evaluate the available KPIs and the parameters. Since terminals have a limited battery capacity, these intensive evaluations are better performed in the network where the power supply is always available.
- The evaluation of available KPIs and parameters can be done more extensive at the network side, for example, through the inclusion of self-learning modules and the access to databases. The possibilities of exploiting available resources to improve the traffic steering decision are much greater at the network side than at the terminal side.
- There are specified mechanisms available to transfer KPIs and parameters from the terminal to the network side such as the IEEE 802.11k for WLANs and the OMA DM DiagMon functionalities supporting all RATs that can gain access to the EPC.

  Hotspot 2.0 can provide network information towards the terminal, but Hotspot 2.0 is limited to the use of Wi-Fi. The ANDSF can provide rather static network information and operator policies to the terminal, such as location specific environment information or static policies.
- The terminal has only a very focused and limited view of its neighbouring environment but no overall view of the surrounding HetNets. Contrary to this, the network side may have a comprehensive overall view of the surrounding HetNets and the operator's view but no specific, terminal-centric view, such as RSS etc. Therefore, the complementary information from each side has to be exchanged in order to make an efficient traffic steering decision.

In this thesis it is proposed that KPIs and parameters are collected at the network side because of the aforementioned reasons. The information available at the terminal is provided to the network to enable a central evaluation point at the network side. This thesis considers a network-controlled, terminal-assisted approach

of traffic steering where the PoD can be either located at the end-user device or at the network. The two supported approaches in this thesis are illustrated in Figure 41. Figure 41 a) shows the case where the PoD is on the terminal side. On the network side the KPIs and parameters are collected and evaluated and a rough decision is made on traffic steering. The policies, which are not strict and therefore allow multiple possibilities, are then sent to the terminal. The terminal evaluates the received policies from the network with the KPIs and parameters at the terminal side and makes the final decision on traffic steering. Figure 41 b) shows the second approach which is supported in this thesis. The PoD is completely located at the network side. Therefore, the policies sent from the network to the terminal are strict and do not allow any uncertainty or margin for the terminal regarding the decision. As a result of these strict policies, the policies act like a command that the terminal has to enforce.

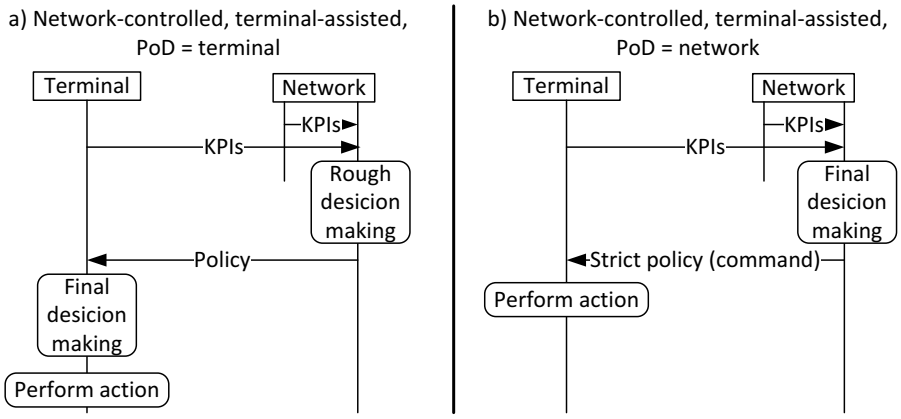


Figure 41: Supported control-modes dependent on the Point of Decision

4.9 Data gathering protocols

KPIs and parameters are required to be integrated into decision making algorithms in order to make appropriate traffic steering decisions. KPIs can be gathered through the OMA DM DiagMon enablers. The OMA DM DiagMon is bearer agnostic, this means the data can be transferred over any RAT. But the DiagMon is only destined to gather information about 3GPP networks and not about WLANs. The Hotspot 2.0 specification allows the end-user device querying the AP for information, but not the other way around. The AP could not query the end-user device for information. Therefore, Hotspot 2.0 cannot be used to gather WLAN related KPIs from the end-user device. To be able to get KPIs from WLANs, the IEEE 802.11k (IEEE 802.11k,

2008) standard can be used instead. Table 11 shows which information gathering technologies can be used to gather information about which RATs.

Table 11: Information gathering technologies per access technology

Gathering Technology	Information about RAT
OMA DM version 2.0	3GPP, WLAN
OMA DiagMon	3GPP, WLAN (only the received power trap in dBm)
IEEE 802.11k	WLAN

4.10 Traffic Steering Management Protocol

This thesis considers a network-controlled, terminal-assisted approach of traffic steering where the PoD is either located in the end-user device or in the network. The two ways of providing policies to the end-user device that finally enforces traffic steering is either with OMA DM or with Hotspot 2.0 release 2. But Hotspot 2.0 release 2 is not yet completely specified, but expected to enable the network side to provide policies to the end-user device.

One requirement for traffic steering management protocols is the capability of delivering policies to the end-user device in a real-time manner in order to ensure that the policies are delivered to the end-user devices in time to enable the end-user device to enforce the policies also in time critical situations, such as 3GPP handover, WLAN roaming, traffic offloading etc. Since the Hotspot 2.0 Release 2 is not yet published, it cannot be verified, whether or not the Hotspot 2.0 Release 2 fulfils the requirement of delivering policies fast enough for time critical situations. In this thesis, the assumption is made that policies can be distributed to the end-user devices using Hotspot 2.0 Release 2.

The OMA DM has evolved continuously from version 1.2, to version 1.3 to the newest version 2.0 released in December 2013. In (Tervonen and Mustajärv, 2010) it has been shown that OMA DM version 1.2 causes problems for real-time applications, since an OMA DM session setup and the transaction of MOs can take several seconds and therefore, the use of the OMA DM version 1.2 protocol is problematic for environments demanding real-time capabilities, such as traffic steering. The newest OMA DM version 2.0 introduces significant improvements compared to the former versions 1.2 and 1.3. The complexity is reduced and the interoperability is improved.

The transaction level of the OMA DM version 2.0 is not backwards compatible due to the use of the new protocol which uses HTTP, RESTful methods and JSON for serialisation. Furthermore, the OMA DM version 2.0 simplifies the access and organisation of the DM tree, where all the MOs are stored, by adding a unique identifier to each MO.

RESTful methods are suitable for real-time applications. Examples are applications in vehicular and sensor networks such as (Ghinamo et al., 2012) (Mizouni et al., 2011) (Jianfeng and Jianglong, 2012). The real-time capability of RESTful style is achieved by the lightweight request/response approach. OMA DM version 2.0 is using JSON for serialisation, which is a very compact format compared to the formerly used SyncML.

In case of a time-critical situation, such as a forthcoming handover, the received signal strength and other observed radio parameters may deteriorate. But this deterioration can be recognised by the traffic steering entity and the DM session can be established very fast to enable a quick provision of the policies to the end-user device via RESTful methods. The policies provided by the traffic steering network entity to the end-user device can either assist the end-user device with the access network selection or in case the policies are restrictive, act as a command and the end-user device´s only option is to enforce the policy. Thus, the PoD is either located at the end-user device or at the network side.

In summary it can be stated, that the primary traffic steering management protocol for this thesis is the OMA DM version 2.0. As the secondary choice, the traffic steering management protocol Hotspot 2.0 Release 2 is considered, keeping in mind that this protocol is able to send policies to the end-user device, but no statements can be made yet about the real-time capability of this protocol.

4.11 Summary

In this chapter, the major design decisions of the thesis have been presented and discussed. The 3GPP RATs are the LTE and LTE-Advanced and Wi-Fi is the non-3GPP RAT for offloading traffic. The selected mobility protocol in the core network, the EPC, as well as for the connection of WLANs to the core network, is the network-based mobility protocol GTP. To enable the connection of the WLANs towards the EPC the S2a interface for trusted WLAN access is selected. To offer solutions for smart,

intelligent and individualised traffic steering policies are selected as the appropriate tool. The IEEE MIHS solution has been ruled out in this thesis, and instead the 3GPP ANDSF approach is selected. In this research traffic steering is performed using the network-controlled, terminal-assisted approach, where the PoD may be either located in the end-user device or in the network. To be able to gather context data from the 3GPP and the WLAN networks, the three protocols OMA DM version 2.0, OMA DiagMon, and the IEEE 802.11k protocol are considered. For the distribution of commands and recommendations to the end-user devices the OMA DM version 2.0 is used as the primary traffic steering management protocol and the Hotspot 2.0 Release 2 is considered as the secondary protocol, because the Hotspot 2.0 Release 2 is not yet standardised completely. The Hotspot 2.0 Release 2 will support the distribution of commands and recommendations, but no statements can be made yet about the real-time capability of this protocol.

5 Black Rider Basic Overview

Recent years have witnessed an exponential growth of traffic in mobile environments, in conjunction with more stringent QoS requirements and seamless integration of heterogeneous access technologies. Unlike bulk data transfer, this traffic growth is likely to be mostly multimedia-based, time-sensitive data, therefore accompanied by increasing requirements on QoS demanded by upper layer services and the always-on-demand from subscribers. Due to the mobility aspect of the end-user devices, these requirements have to be fulfilled even when the end-user device is moving. In order to satisfy these demands and the predicted future growth, MNOs must enhance their infrastructure to allow appropriate management and distribution of the traffic through providing for example smart handover decisions and traffic offloading onto other than 4G physical networks like Wi-Fi, WiMAX, and future network technologies. The predicted exponential growth and the fulfilment of the QoS and mobility requirements must be accompanied in the MNOs infrastructure, by enabling an appropriate traffic steering strategy for HetNets. The BR introduces a new architecture including the new network element the BR enabling generic context- and policy-based traffic steering in heterogeneous wireless networks. The BR provides individual traffic steering functions on an end-user device level. Traffic steering is an important factor for QoS fulfilment and mobility, because it influences the QoS relevant parameters such as the available bandwidth, the delay, the jitter, and the packet loss while steering the end-user device through the appropriate and available access networks. The design and implementation of an intelligent traffic steering system requires considering a number of impact parameters from different stakeholders, such as the subscriber, the MNOs, and the upper layer services. But, beside the requirements of the stakeholders, the capabilities of the network nodes and the availability of the access networks have to be considered as well. In order to be able to fulfil the requirements and take into account the capabilities of all the involved network components, context information is a key factor for managing such a complex environment. Context information has also been used in traditional cellular networks such as GSM, GPRS and UMTS. This context information, for example, has been constituted by system information distributed on the Broadcast Control Channel

(BCCH). Additionally, policies have been applied to define, for example, under which conditions a UE is allowed to change the cell.

The proposed BR architecture extends the common control information, sent over the broadcast channel, with individual control information and policies per end-user device. This individual control information is provided by the BR to the end-user devices in a unicast manner.

To be able to provide intelligent and efficient traffic steering, appropriate information from the different stakeholders (end-user devices and MNO networks) has to be gathered in a real-time manner, because some information is only useful if it is up-to-date. For example, information, such as capacity and utilisation of the core network and the radio cells or the location of an end-user device are of significant importance in order to guide the end-user device through the available networks by making appropriate handover or offloading decisions. However, the information is valid and valuable for traffic management purposes only if it is current information. To be able to fulfil the requirements of the involved network entities, the BR gathers information from the end-user device and from the network elements in order to manage this complex environment.

Traditional cellular network handover control as well as the PoD are located at the network side. The ANDSF assists the end-user device in the network selection decision only by means of previously assigned network selection policies and not via real-time information evaluation. In contrast to the ANDSF the BR decisions are based on real-time information, because the BR is able to gather actual context data and use it to create policies that reflect the current situation of the involved stakeholders. The BR supports a network-controlled, terminal-assisted approach of traffic steering with either the PoD located at the terminal or at the network, as defined in section 4.8. As a result, in the network-controlled mode with the PoD located at the terminal, the BR sends just recommendations to the end-user device without making any final decision. In the network-controlled mode with the PoD located at the network the BR makes the final traffic steering decision and sends the result as a strict policy, which acts like a command, to the end-user device.

5.1 Black Rider Capabilities

In the following, two main capabilities are briefly introduced to provide a high level view of the BR.

5.1.1 Individual Traffic Steering

The BR provides the capability of individualised traffic steering. As a result, commands and recommendations are individually created per end-user device. Due to the individually created commands and recommendations per end-user device the BR has the capability to take into account the individual parameters and KPIs, such as the user preference, the location, the speed, the available interfaces at the end-user device, the available networks etc. to incorporate them into the traffic steering decision. The generated commands and recommendations define or limit the handover and offloading behaviour individually per end-user device and therefore allowing for an individual traffic steering solution considering the individual QoS requirements and mobility behaviour. The individually generated commands and recommendations per end-user device help optimising the traffic load essentially at the access networks by considering requirements of different stakeholders.

5.1.2 Provide Information to 3rd Party

The BR is able to provide information to 3rd parties. Within this research accurate geolocation information is used as an example of information provided to 3rd party. Geolocation information is used for several services, such as Content Delivery Networks (CDNs), cloud balancing, direct marketing, context-sensitive content delivery etc. Accurate geolocation information could be used by CDNs to improve the accuracy of their services. CDNs are widely used to reduce the delay by applying IP geolocation service to identify and select a close-by server to deliver the requested content. As a result of the shorter path, the delay is reduced. A common technique for realising IP geolocation uses databases that map IP address blocks to geographical locations. Another technique is the active way of getting geolocation information, by measuring the delay. With this technique, more accurate geolocation information can be achieved, but with the lack of scalability, high measurement overhead, and very high response time ranging from tens of seconds to several minutes to localise a single IP address (Poese et al., 2011). Geolocation services are only as valuable as the

geolocation information is accurate. In (Poese et al., 2011) several geolocation databases – both, commercial and free accessible – have been analysed and compared to gain insights into the limitations in their usability. The accuracy of geolocation databases with a large European Internet Service Provider (ISP) has been quantified based on ground truth information. The term ground truth refers in this case to the comparison of predicted location with the real location. The conclusion has been that the geolocation databases can claim country-level accuracy, but not city-level accuracy. In contrast, the BR concept could provide exact city, zip code and even cell accuracy for end-user devices which are connected over the EPS and are active. As a result, the exact geolocation information, provided by the BR, enables more accurate services which are based on this geolocation information.

5.2 Enablers for the Black Rider Concept

The enablers for the proposed new context- and policy-based traffic steering architecture for heterogeneous wireless networks are the functional BR network element, the common database solution in the EPS, the information gathering mechanisms, and a traffic steering management protocol for delivering commands and recommendations from the BR to the end-user devices. These elements are briefly introduced in the following.

5.2.1 The Black Rider

The BR is a functional network element providing intelligent, generic context- and policy-based traffic steering for HetNets, which is applicable for a broad spectrum of use cases and optimisations. The BR is responsible for the up-to-datedness of the context and policies of every end-user device which is in the serving area of the BR. External modules are operator defined or 3rd party defined modules that can be applied to make appropriate decisions on traffic steering. External modules are designed to analyse or manage special data and, as a result, generate predictions, recommendations or decisions. Examples of external modules are mobility analysers and handover or offloading managers. The BR acts as a coordinator for external modules. External modules can be concatenated by the BR to improve the overall traffic steering decision. The commands and recommendations are distributed by the BR using a traffic steering management protocol to the end-user device supporting a

network-controlled mode either with the PoD at the terminal or network side, as described in section 4.8. As a result of this, either the BR or the end-user device is in charge of the final decision making process. The BR can provide individual traffic steering per end-user device.

5.2.2 The Common Data Base

Up to 3GPP Release 8, all the information is stored in a distributed manner in several network entities of the EPS. For example, beside the HSS database, nearly every network entity has its own, small database for permanent data, for example, the MME, the SGW, the PDN GW, and the eNB. Network entities, which hold permanent information support several different protocols and as a result of this, it is very difficult and takes a huge effort to collect all the necessary data from each of the network entities. Since 3GPP Release 9, a database solution, the User Data Convergence (UDC) (3GPP TS 23.335 V10.0.0, 2011), has been defined with the aim of reducing the amount of entities, which are holding databases and to replace these mini databases storing permanent data by introducing one common and logically unique database, the User Data Repository (UDR). Figure 42 shows the UDC reference architecture. The key feature of the UDC concept is to separate the application logic from data and to harmonise data access for different protocols. Therefore, different application Front Ends (FEs) are introduced. FEs have the function of mediators between the multiple reference points based on different protocols and the UDR interface Ud. The supported protocols at the network interfaces are not limited to them shown in Figure 42. If a new protocol/reference point is added, an appropriate FE needs to be added to the UDR. The Ud interface stack for data access is LDAP/TCP/IP/L2/L1 and for subscription notification it is SOAP/HTTP or HTTPS/TCP/IP/L2/L1. One FE can serve one or multiple network elements.

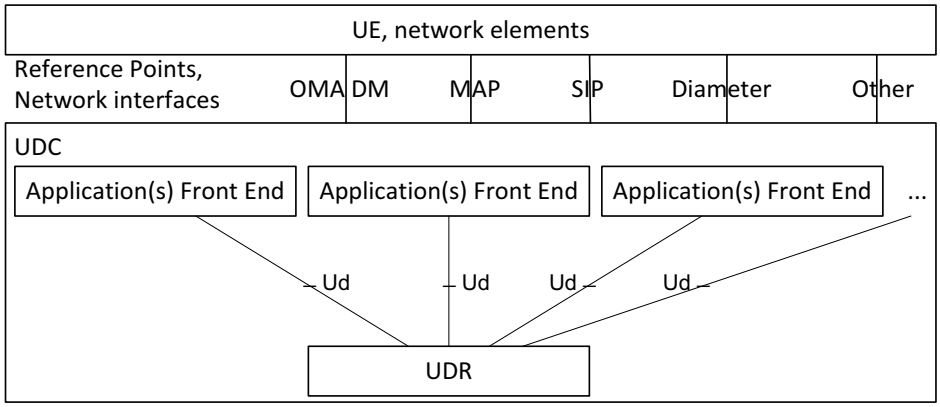


Figure 42: UDC reference architecture

The UDR includes a subscription and notification service to notify interested entities of relevant changes. To avoid unauthorised access, an access control subsystem is included in the UDR. Therefore, the UDR is appropriate for 3rd party access, because the access can be controlled. The Create, Read, Update, and Delete (CRUD) operations are available at the UDR.

By applying the UDC concept, the BR can collect its required data from one common database and can gather individual and personalised contexts on a per end-user device basis. The network entities can access the UDR by using the harmonised access through the Ud interface. There are user data categories defined in the UDR standard (3GPP TR 22.985 V10.0.0, 2011) that are not typically subject of the UDR, such as user content data (containing e.g. application related data such as photos or videos), or user behaviour data (consisting of call- or session-related dynamic data, registration status etc.). The BR provides data gathering mechanisms to also collect these valuable data categories.

5.2.3 Information Gathering Mechanisms

KPIs are of paramount importance for an appropriate traffic steering. Therefore, the BR needs information gathering mechanisms to collect information about the capabilities, states, locations etc. of the involved stakeholders to be able to guide the end-user devices through the heterogeneous mobile networks and determine the most suitable and appropriate access network. Core network entities and end-user devices can act like sensors, being able to provide information about KPIs to the BR. This information can be stored at the BR UDR. Once the information is available to the BR, the traffic steering decisions can be made on basis of the gathered KPIs.

5.2.4 Traffic Steering Management Protocol

The traffic steering management protocol is responsible for delivering commands and recommendations in a real-time manner to the end-user devices. The distribution of the commands and recommendations has to be fast enough to handle time-critical situations, such as handover. The selection of the traffic steering management protocol has already been made in section 4.10. The OMA DM version 2.0 is selected as the primary traffic steering management protocol for this research. As the secondary traffic steering management protocol the Hotspot 2.0 Release 2 is considered, with the fact in mind that this protocol is definitely able to send policies to the end-user device, but no statements can be made yet about the real-time capability of this protocol.

5.3 Summary

The complex environment, the challenges and demands from the end-user and the network side have been explained. Reasons for the need of an overall concept to steer traffic in a smart and intelligent, as well as individualised way to tackle these upcoming problems of high traffic load on the mobile networks have been outlined. The BR has been introduced by giving a high level view of the main capabilities of the BR. Furthermore, the four enablers for the BR architecture are explained: The BR itself, the need for a common database, mechanisms to gather information to collect the individualised end-user device contexts, and the traffic steering management protocol. All these enablers have been described on a high level to provide an entry point in order to get deeper into details in the following chapter 6.

6 Black Rider Architecture

The goal of the BR architecture is to enable individualised traffic steering per end-user device taking into account user and network context data. The architecture contains all the required building blocks of the surrounding environment of the BR. A brief overview of the involved components is depicted in Figure 43 including the surrounding layers. On top there is the service layer consisting of, for example, the IMS and the Next Generation Service Overlay Networks (NGSONs). The service layer has access to the common database, the UDR. 3rd parties can access information gathered by the BR. The BR contains a BR database (DB), which is a modified UDR with no restrictions on data types. It is allowed to store any valuable information in the BR DB. The 3GPP and non-3GPP access networks, as well as the end-user devices and any other network entities act like sensors, because they can provide context data and in the case of end-user devices also policies to the UDR and to the BR. The BR itself is located in a cloud.

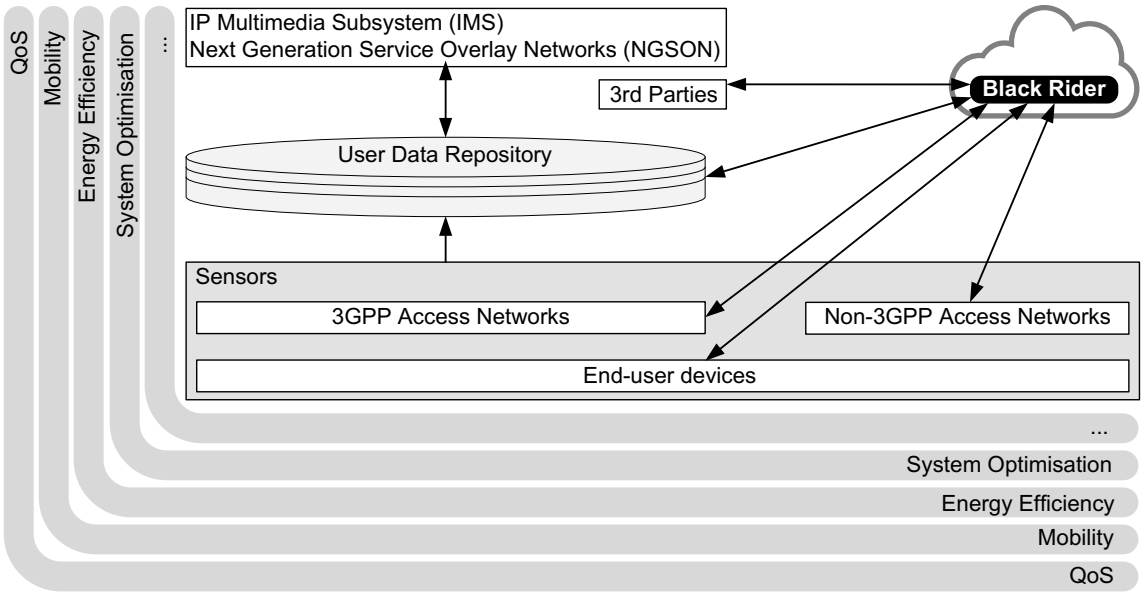
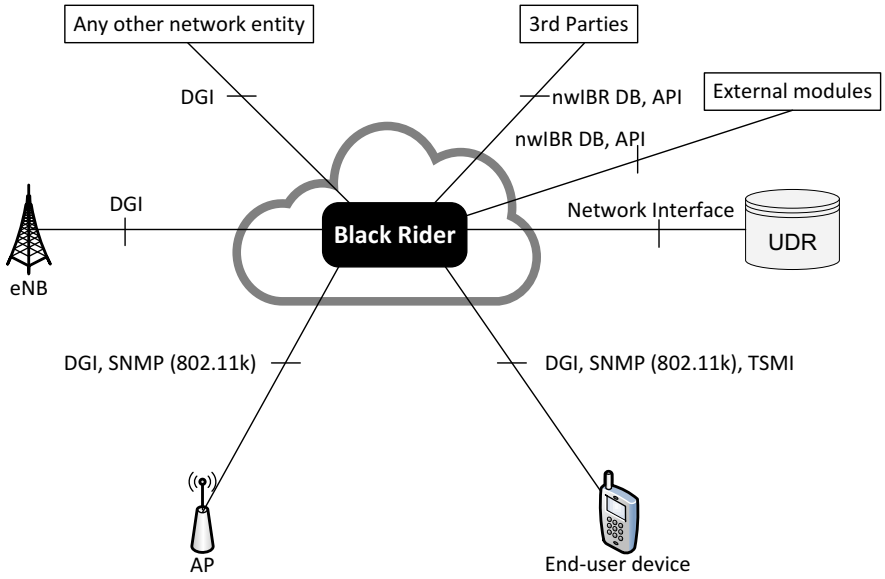


Figure 43: Brief overview of the Black Rider architecture

The BR architecture enables optimisations in the area of mobility and QoS management, energy efficiency, and information provisioning to 3rd parties. Possible use cases are provided QoS, the choice of the appropriate access technology, handover decisions, offloading decisions, energy efficiency, CDNs, marketing related

use cases etc. These improvements are based on the fact, that the BR manages and exploits the user contexts and the policies per end-user device in an individual way. Whenever an end-user device is switched on, the BR starts gathering actual context and policy data through the Data Gathering Interfaces (DGI). The DGI are specific interfaces that are provided by the BR architecture shown in Figure 44. The SNMP (IEEE 802.11k, 2008) is used for data gathering from WLAN APs and WLAN capable UEs described in section 3.4.5. The BR is located at the cloud and provides multiple interfaces and protocols for network entities to access the BR. The interface between the BR and the UDR is called network interface. As explained in section 5.2.2, different protocols are allowed to run on the network interface between the application FE, providing access to the UDR, and the network entity. The same interface is used to provide external modules access to the BR DB, which is a modified UDR. But to differentiate these interfaces, the interface between the BR and the external modules is called network interface BR DB (nwI_BR DB). The Application Programming Interfaces (APIs) between the external modules and the BR provide the external modules access towards the BR. The APIs can be adapted and extended according to the needs of the external modules and the mobile operator guidelines. The APIs between 3rd parties and the BR are used to encapsulate specific BR functions to expose them to 3rd parties. The Traffic Steering Management Interfaces (TSMI) are used to deliver the commands and recommendations to the end-user devices.



Data Gathering Interfaces **(DGI):** nwIBR DB, OMA DM v2.0 (REST), OMA DiagMon v1.2
Traffic Steering Management Interfaces **(TSMI):** OMA DM v2.0 (REST), Hotspot 2.0

Figure 44: Black Rider interfaces

6.1 Black Rider Building Blocks

The BR consists of three functional building blocks shown in Figure 45.

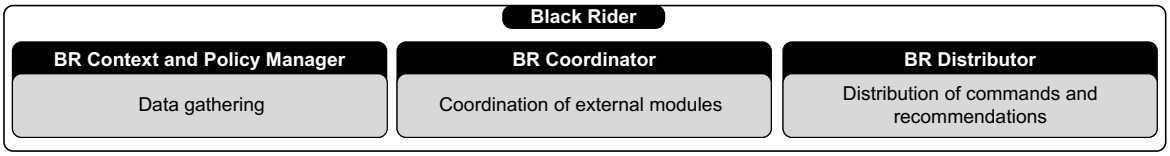
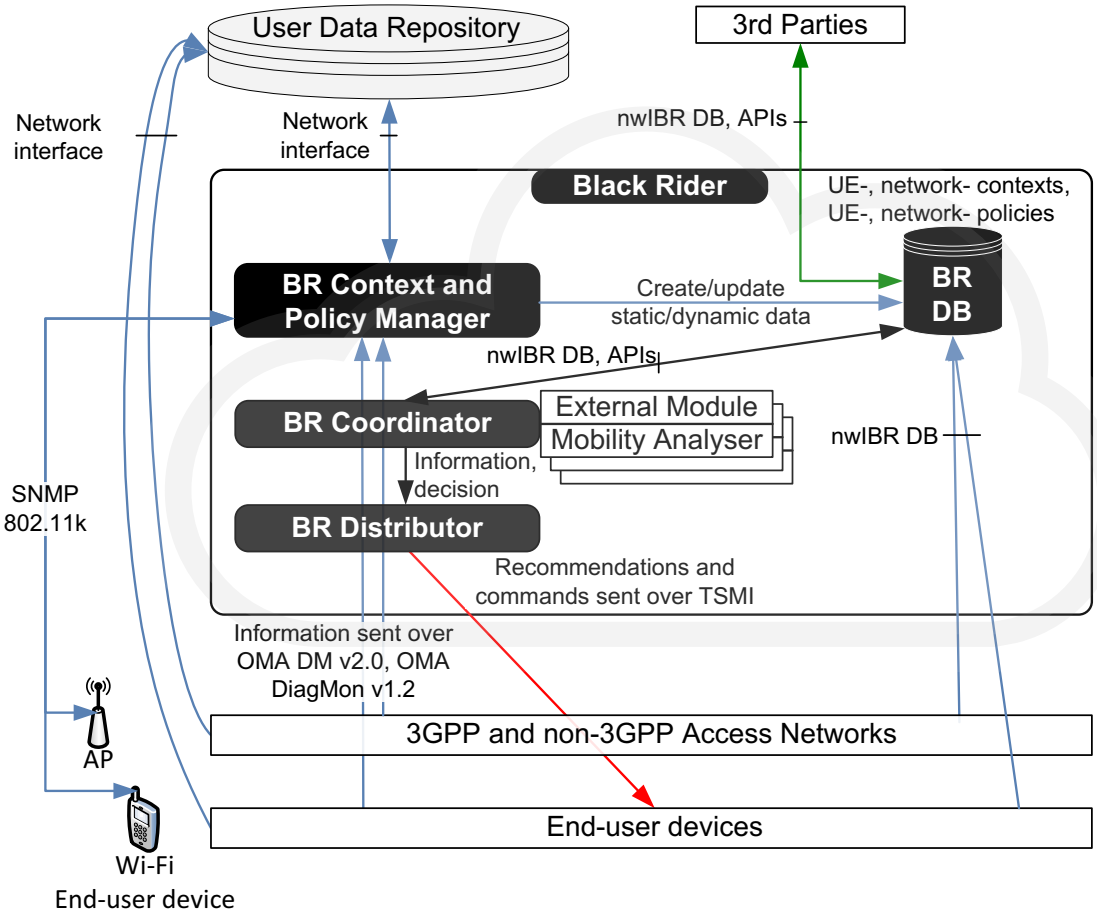


Figure 45: Functional building blocks of the Black Rider

- **BR Context and Policy Manager:** performs data gathering. The data gathering of static and dynamic real-time data of the MNO´s network, and the end-user devices is the basis for establishing the user specific context and policies which can be used to make smart and appropriate traffic steering decisions for a specific end-user device.
- **BR Coordinator:** performs the coordination of external modules. The external modules, such as mobility analysers (Gajic et al., 2012), (Jeong et al., 2011), (Chang et al., 2009), vertical and horizontal handover management (Gondara and Kadam, 2011), offloading management etc. could be triggered and concatenated by the BR to achieve an appropriate traffic steering decision for a particular end-user device.
- **BR Distributor:** distributes the commands and recommendations. Once a decision is made by the BR, the command(s) or recommendation(s) have to be sent to the end-user device, where the decision is enforced or the recommendations are processed.

Figure 46 provides a more detailed view of the BR architecture showing the BR building blocks. The blue lines indicates data gathering mechanisms, while the green line indicates the data provision to 3rd parties and the red line indicates the traffic steering commands and recommendations by the use of the TSMI containing the OMA DM v2.0 (REST) and the Hotspot 2.0 Release 2 protocol. The black lines indicate internal communication of the BR. The BR Context and Policy Manager is responsible for an updated and persistent context and policies for the specific end-user device assigned to the BR. Information can be retrieved from the UE, 3GPP and non-3GPP accesses by the BR Context and Policy Manager either via the OMA DM version 2.0 protocol in combination with the DiagMon version 1.2 or in a direct manner from the network entities or end-user devices via the nwIBR DB interface by the BR DB. For

WLAN APs and end-user devices the information is gathered via the SNMP IEEE 802.11k standard. The BR Context and Policy Manager can subscribe to certain events at the UDR and be notified by the UDR, for example if a certain threshold is exceeded, or a certain data set is changed. As a result of such a notification through the OMA DM version 2.0, the DiagMon version 1.2 or the SNMP IEEE 802.11k standard, the BR Context and Policy Manager can load the current information into the BR DB.



Traffic Steering Management Interfaces **(TSMI):** OMA DM v2.0 (REST), Hotspot 2.0 Release 2
Blue: Data gathering, **Red:** Traffic steering,
Green: Data provisioning, **Black:** Internal BR communication

Figure 46: Detailed Black Rider architecture

The current information in the BR DB can be accessed and processed by the BR Coordinator to make, for example, the most appropriate choice of the access network, handover decisions or to offload traffic, according to offload strategies for each end-user device in a personal and individual manner. The BR Coordinator can trigger external modules for decision making and provide current and required information to them. The BR considers the requirements and status information of the involved network entities, the end-user devices, as well as the policies, and provides this

information to the external modules. An example of an external module is an external mobility analyser evaluating the mobility behaviour of the user and identifying user movement habits, such as the identification of stable locations like workplace, university, home and usual transits from one stable location to another. Based on these findings, a prediction of the next movement of the user can be made. There are a lot of proposals for mobility analysers that can be used by the BR, such as described in (Gajic et al., 2012), (Jeong et al., 2011), and (Chang et al., 2009). To achieve a more accurate handover decision and access network selection including, for example, also offloading strategies, the BR Coordinator may concatenate the services offered by multiple external modules.

For example, after the BR Coordinator has received the result from a mobility analyser, the result and other relevant information may be handed over to another external module which analyses the speed of the end-user device and evaluates the surrounding cell arrangements to finally select the most suitable cell and cell size for this end-user device. The decision may contain the selection of an umbrella or macro cell size, if the speed of the end-user device exceeds a certain speed threshold, in order to prevent the end-user device from performing frequent handovers from one small cell to another. The BR can improve the consequent and intelligent enforcement of demanded requirements by taking into account the provided capabilities of the stakeholders. If the BR Coordinator has a result in form of either a decision or recommendation(s) it triggers the BR Distributor which generates and sends the command(s) or recommendation(s) using the traffic steering management protocol to the appropriate end-user device. The BR also provides a service enabler interface for 3rd parties. A service enabler interface provides the ability of enabling new services to 3rd parties, based on the provided information by the BR, such as the provision of geolocation information to a 3rd party. This use case is discussed in more detail in section 7.3.

6.2 Data Gathering

The real-time data gathering of KPIs and parameters is significant in enabling an individual and appropriate traffic steering decision making for a particular end-user device. The BR is responsible for gathering an individual and personalised context and a current set of policies.

6.2.1 Data Classification

Table 12 shows the different kind of data classes and associated changing rates: rarely, dynamic, highly dynamic or hybrid, which is a combination of the previously named changing rates. The data classes are: the end-user device context, the end-user device policies, the network context, and the network policies. The end-user device context contains data that is related with the end-user device or the end-user. This includes all information collected through the end-user device which is end-user device related, as well as the subscription information of the end-user. The end-user device policies include user preferences, operator defined policies and operator defined policies that are chosen by the user or automatically. The latter is a new kind of policy introduced in this research which is operator defined but the user can choose the policy or the policy is selected automatically. An example of this kind of policies is the energy model. The energy models can for example consist of the following operator defined models: Low Energy Consumption (LEC), Medium Energy Consumption (MEC), High Energy Consumption (HEC), Automated (AUT). The selection of the energy model has not only an impact on the energy consumption of the end-user device but also on the QoS, because the energy model selection may affect the selection of the RAT and different RATs usually have different QoS characteristics. As a result, the selection on the RATs is influenced by the selected energy model, because different RATs usually have different energy consumption levels. Either the user itself can select the energy model or the selection is performed automatically by the BR. The operator defined policies include usually predefined thresholds and rules that have to be complied with. The user preferences are rules that are defined and selected by the user. The network context consists of general network information, such as the kind and location of RATs, cell locations, used frequencies, cell identifier etc. This kind of information rarely changes. In comparison to this stable information there is the specific network context information which consists of dynamic to highly dynamic information about the state of the network, such as the antenna utilisation, the current network load, the throughput etc. The possibility of exploiting this kind of data for traffic steering is new and not supported so far within the 3GPPs defined ANDSF. The network policies contain the operator defined policies, such as restricted RATs and PLMNs, RAT priorities etc.

Table 12: Data classification

END-USER DEVICE CONTEXT		
Kind of data	**Description**	**Changing rate**
Subscription information	Such as basic service, value added service, class, e.g. VIP etc.	Rarely
Device Management	End-user device information, such as battery charge level, available interfaces, device capabilities etc.	Rarely to dynamic
Session Management	Bearer information, such as numbers and kind of bearer, EPS QoS value per bearer etc.	Dynamic
Mobility Management	Such as current location, history, RAT, measurement reports etc.	Dynamic to highly dynamic
Collected information	Such as discovered home zone, RAT frequencies etc.	Dynamic
END-USER DEVICE POLICIES		
Kind of data	**Description**	**Changing rate**
Operator defined policies - chosen by the user or automatically	Such as Energy model, cost model (cheapest solution, medium solution, highest Quality) etc. This user defined policies are basically pre-defined operator models where the user can choose of.	Rarely to dynamic
User preferences	Preferred RAT, available interface(s), maximum cost etc.	Rarely to dynamic
NETWORK (NW) CONTEXT		
Kind of data	**Description**	**Changing rate**
General NW information	The kind and location of RATs, the location of cells, the used frequencies, BSSID, SSID, cell ID, bandwidths etc.	Rarely
Specific NW information	The utilisation of the antennas, network load, throughput etc.	Dynamic to highly dynamic
NETWORK (NW) POLICIES		
Kind of data	**Description**	**Changing rate**
Operator defined policies	Operator defined policies such as restricted access technologies, restricted PLMNs, priorities of RATs etc.	Rarely to dynamic

6.2.2 Data Gathering Variants

Figure 47 shows the different interactions over different interfaces between network entities and the BR to either enable the BR to gather information such as KPIs or provide information to 3rd parties. The BR Context and Policy manager is able to gather data via the OMA DM server and DiagMon server, the SNMP IEEE 802.11k module, the BR FE, and the BR DB. 3rd parties can access data in a direct manner at the BR DB via FEs or via APIs.

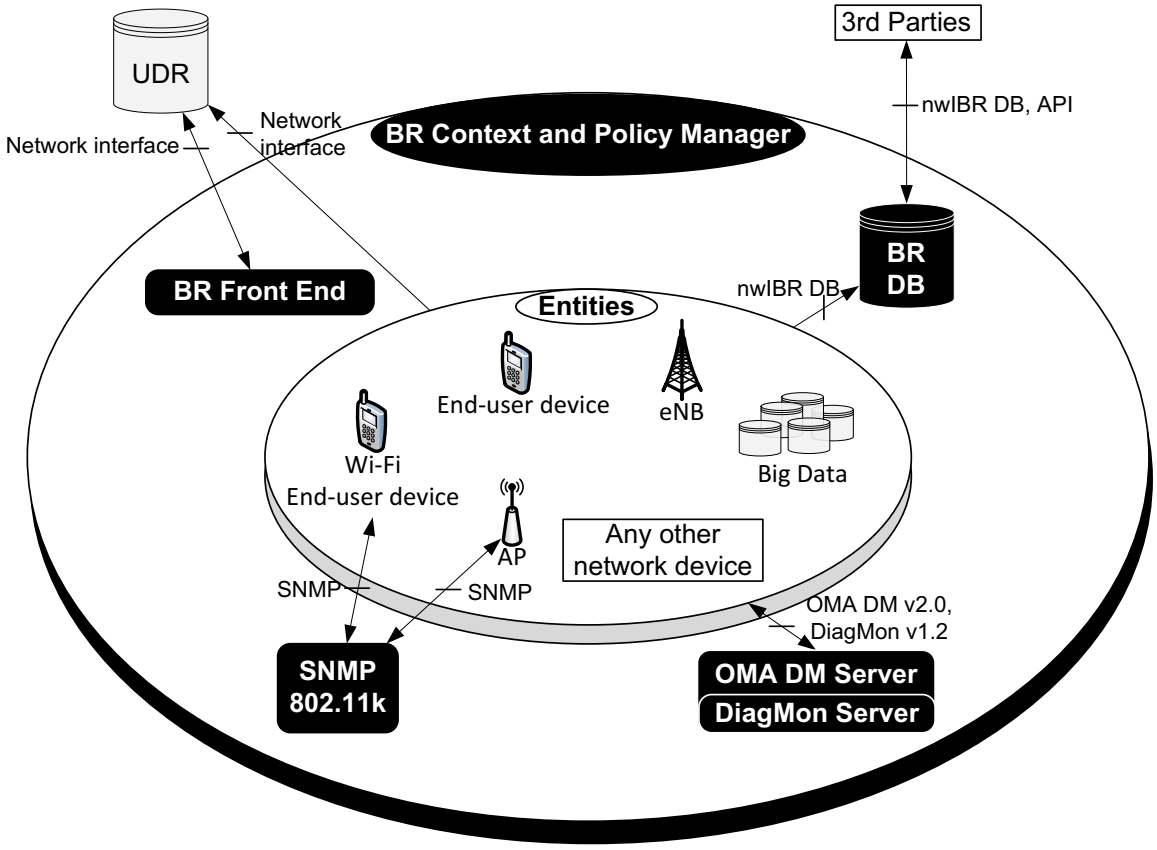


Figure 47: Overview of the Black Rider data gathering variants

For contexts and policies that rarely change, it is sufficient to gather data in a passive manner. This means that network entities trigger autonomously the creation of new data sets or the update of existing data sets whenever needed. Variants of passive data gathering procedures are shown in sequence 1 to 5 of Figure 48. The passive data gathering provides no opportunity to the BR to actively request data from any network entities. Hence, the BR is dependent on the information provided by the network entities to the BR. Sequence 4 of Figure 48 provides a more current way of data gathering compared to the sequences 1 to 3 and 5 of Figure 48, because it uses the trap mechanism of the OMA DiagMon. The trap mechanism enables the BR to set thresholds and be notified if these thresholds are exceeded or fall below. For dynamic or highly dynamic changing rates of contexts and policies, passive data gathering procedures are usually not current enough for time-critical processes, such as a handover. The only exception is the applied trap mechanism shown in sequence 4 of Figure 48. Therefore, the BR requires further the possibility of gathering data in an active way by explicitly requesting specific information from the network entities. This is shown in sequence 6 to 8 of Figure 48.

The sequences 1 to 3 of Figure 48 use the UDR and the BR DB-based concept to gather data from the end-user device. Sequence 4 of Figure 48 uses the OMA DiagMon trap framework. The BR Context and Policy Manager implements a DiagMon server, thus the BR Context and Policy Manager can register on traps and be notified if a trap becomes active or inactive. Sequence 3 of Figure 48 is the most direct passive data gathering method, because the end-user device updates or creates data sets directly in the BR DB. In all other cases, the BR Context and Policy Manager has to update or create data sets in the BR DB as a consequence of the received new or changed values. Sequences 5 and 6 of Figure 48 use the OMA DM version 2.0 applying the REST architectural style (Fielding, 2000). The REST architectural style allows several other, enhanced scenarios, for example the use of proxy servers. It is possible to integrate these enhanced scenarios of the REST architectural style into the BR architecture, but in this thesis only the basic use cases of REST are considered to enable network entities to provide information to the BR (sequence 5 of Figure 48) and to enable the BR to actively request information with REST architectural style (sequence 6 of Figure 48). Since the OMA DiagMon can be used also with the latest OMA DM version 2.0 using REST architectural style, sequence 6 of Figure 48 does not only represent OMA DM REST architectural style, but also OMA DiagMon data gathering using OMA DM REST architectural style. Sequence 7 and 8 of Figure 48 use the SNMP data gathering mechanism defined in the IEEE 802.11k (IEEE 802.11k, 2008) amendment to actively query Wi-Fi devices such as APs and STAs.

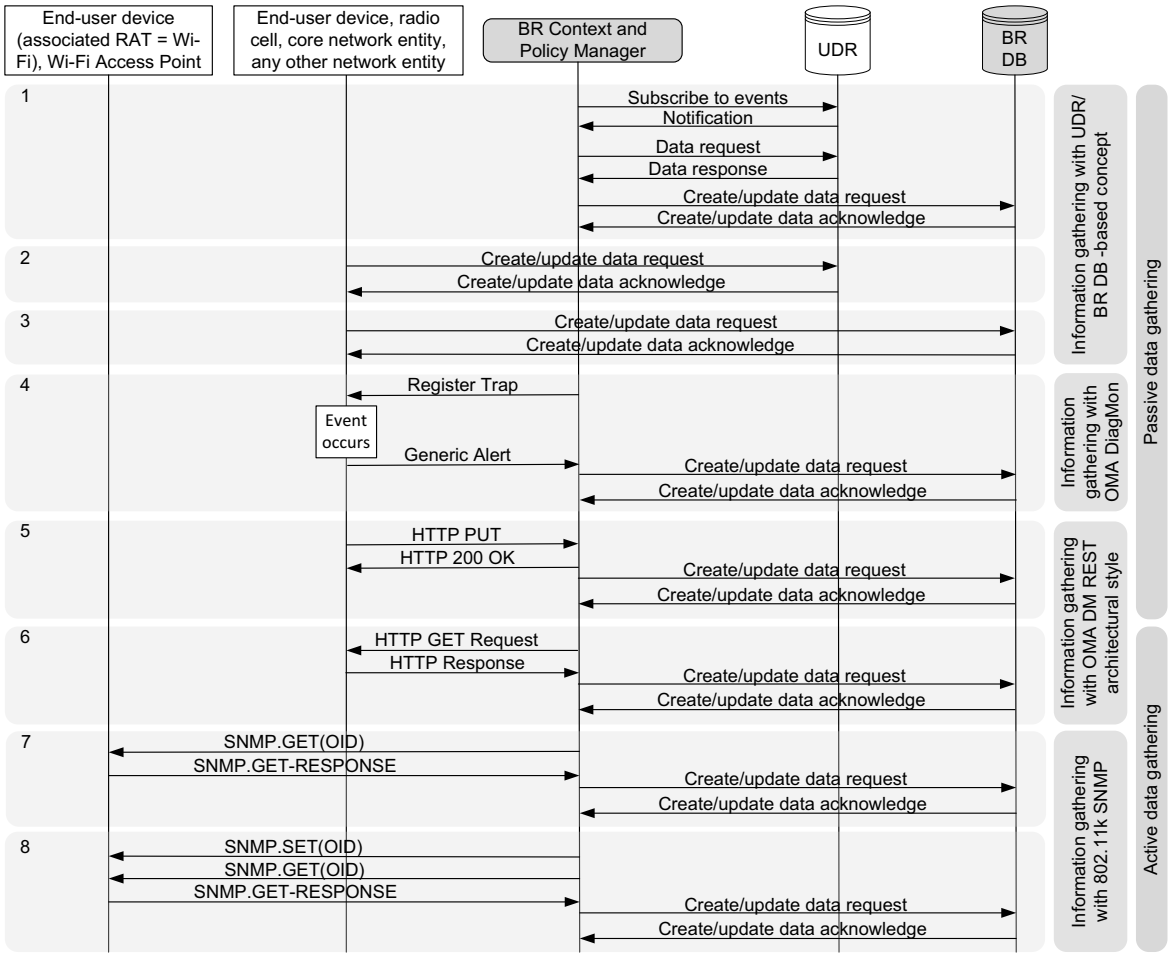


Figure 48: Sequence diagram of data gathering variants

The BR Context and Policy Manager is responsible for keeping the required data from the UDR or any other network entity updated in the BR DB and requesting required information from network entities. In the following, the different variants of data gathering, depicted in Figure 48, are explained in more detail.

1. The BR Context and Policy Manager can subscribe to events at the UDR to be notified if a certain parameter, stored in the UDR, exceeds or falls below a defined threshold. If such an event occurs, the BR Context and Policy Manager is notified by the UDR and can request specific data. If information is received by the BR Context and Policy Manager it has to update the BR DB, either by creating a new- or updating an existing data set.
2. End-user devices, RAN entities, core network entities or any other network entity, can act as sensors. Therefore, they can either create a new- or update an existing data set at the UDR. The restrictions of the UDR are in effect if these access cases are used and therefore storing of content and behaviour data is

not possible. Whenever an updated or created data set at the UDR exceeds or falls below a defined threshold set in sequence 1. the data gathering procedure from sequence 1. is triggered.

3. As in sequence 2 from Figure 48 the end-user devices, the RAN entities, the core network entities or any other network entity, can act as sensors. The network entities can either create a new- or update an existing data set directly at the BR DB. The restrictions of the UDR are not in effect if these access cases are used and therefore storing of the content and behaviour data is possible.
4. The OMA DiagMon provides the capability of using traps. Traps are similar to events that can be raised at the end-user device. A trap contains a threshold which triggers an action if the value exceeds or falls below the threshold. The DiagMon server which is represented in this research by the BR Context and Policy Manager, can register itself at the end-user device for one or multiple traps that is/are supported by the end-user device. If a trap is activated or deactivated at the client the DiagMon server is notified with a generic alert message by the end-user device.
5. End-user devices, RAN entities, core network entities or any other network entity, can act as sensors also with REST architectural style using HTTP. The network entities can send information to the BR Context and Policy Manager which either create a new- or update an existing data set at the BR DB.
6. The BR Context and Policy Manager is able to request required data from the end-user devices, from the RAN entities, from the core network entities or from any other network device. To achieve this, REST is used as the architectural style and HTTP is applied as the protocol. If the requested data is received at the BR Context and Policy Manager it has to update the BR DB, either by creating a new data set or updating an existing data set.
7. The BR Context and Policy Manager can request required data from Wi-Fi capable end-user devices which are associated with a Wi-Fi access network or from a Wi-Fi AP. Whenever possible, the data gathering request should be destined for the AP and not the Wi-Fi capable end-user device STA. As a result of this, the air link is not utilised with management data more than necessary. Data can be queried by the BR Context and Policy Manager through SNMP

requests/responses. These SNMP mechanisms are defined in the IEEE 802.11k amendment (IEEE 802.11k, 2008). To get the appropriate data, the SNMP request has to contain the specific Object Identifier (OID).

8. The BR Context and Policy Manager can request the STA, which is associated with a Wi-Fi access network, to perform measurements on the physical or MAC layer with an SNMP.SET command including the specific OID. Afterwards, the measured results can be requested by the BR Context and Policy Manager with the specific SNMP request and OID.

Beside these named data gathering variants another one exists that can be used as an information source for the BR. Probes systems are used to monitor the network. Probes systems are able to capture messages on interfaces. This is used to detect errors in the network and incorrect behaviour taking into account the end-to-end view. This information source can deliver big data about the network. The gathering variants shown in Figure 48 are designed for specific information gathering from end-user device or single network devices. The KPIs are gathered in a direct and specific manner from the single network entities. In contrast, big data is useless unless it is analysed and processed to extract the desired network KPIs, because the gathered data volume is huge. The BR itself is not designed to process big data. But the results of the processed big data can be used by the BR to include them for the decision making process.

6.2.3 Gathered Key Performance Indicators

In order to gather the personalised and individual context and policies of a particular end-user device, KPIs of different stakeholders are collected to enable real-time context- and policy-based traffic steering in heterogeneous wireless networks. In the following, the specific KPIs per data gathering method are listed.

The data gathering in Wi-Fi access networks can be performed using the IEEE 802.11k amendment. The request/report pairs are defined in the standard. But not all request/report pairs are appropriate for data gathering for the BR. The BR can gather information from both IEEE 802.11 network elements, the STA and the AP by querying the reports from the AP or STA via SNMP.GET messages. The query of the AP instead of STAs has the advantage that the air link is not utilised by SNMP queries. This method is the favoured one. Therefore, the Table 13 provides a summary of

which kinds of measurement results of the request/report pairs are used by the BR to gather data. In principle, all the request/report pairs from Table 13 can be used by the BR, but the ones highlighted with orange colour causes additional utilisation to the air link, and since the air link is usually a bottleneck for transmission, the utilisation with additional management data should be minimised. Therefore, whenever possible, these two orange request/response pairs in Table 13 should not be used by the BR.

Table 13: Accessible IEEE 802.11k reports for the Black Rider

Request/Report pair	Accessible for Black Rider
Beacon	Yes
Frame	Yes
Channel Load	Yes
Noise Histogram	Yes
STA Statistics	Yes
Location Configuration	Yes
Neighbour Report	Yes
Link Measurement	No, but the Received Channel Power Indicator (RCPI) and the RSNI values are available through the Frame Request/Report pair. The BR can access this data with SNMP Get requests directly from the STAs, but in order to not load the air link with additional signalling effort, the BR does not gather Wi-Fi KPIs directly from STAs whenever possible.
Transmit Stream/Category Measurement	No. The transmit stream/category measurements are only available from STAs and not from APs. The BR can access this data with SNMP Get requests directly from the STAs, but in order to not load the air link with additional signalling effort, the BR does not gather Wi-Fi KPIs directly from STAs whenever possible.

Table 14 shows the KPIs per report. These KPI values can be accessed by the BR through SNMP.GET requests. On the other hand it is possible for the BR to initiate measurements at the STAs with the SNMP.SET request.

Table 14: Key Performance Indicators per IEEE 802.11k report

Report	Key Performance Indicators
Beacon	MeasuringSTAAddr, ChanNumber, RegulatoryClass, ActualStartTime, MeasurementDuration, PhyType, ReportedFrameType, RCPI, RSNI, BSSID, AntennaID, Parent Timing Synchronisation Function (TSF), ReportedFrameBody
Frame	ChanNumber, RegulatoryClass, ActualStartTime, MeasurementDuration, TransmitSTAAddress, BSSID, PhyType, Avg Received Channel Power Indicator (RCPI), RSNI, LastRCPI, AntennaID, NumberFrames
Channel Load	MeasuringSTAAddr, ChanNumber, RegulatoryClass, ActualStartTime, MeasurementDuration, ChannelLoad
Noise Histogram	ChanNumber, RegulatoryClass, ActualStartTime, MeasurementDuration, AntennaID, Average Noise Power Indicator (ANPI), Idle Power Indication (IPI) IPIDensity0, IPIDensity1, IPIDensity2, IPIDensity3, IPIDensity4, IPIDensity5, IPIDensity6, IPIDensity7, IPIDensity8, IPIDensity9, IPIDensity10

Statistics	STAAddress, MeasurementDuration, GroupID, TransmittedFragmentCount, MulticastTransmittedFrameCount, FailedCount, RetryCount, MultipleRetryCount, FrameDuplicateCount, Request to Send (RTS) SuccessCount, RTSFailureCount, Acknowledgement (ACK) FailureCount, QosTransmittedFragmentCount, QosFailedCount, QosRetryCount, QosMultipleRetryCount, QosFrameDuplicateCount, QosRTSSuccessCount, QosRTSFailureCount, QosACKFailureCount, QosReceivedFragmentCount, QosTransmittedFrameCount, QosDiscardedFrameCount, Qos Media Access Control Protocol Data Units (MPDUs) ReceivedCount, QosRetriesReceivedCount, ReceivedFragmentCount, MulticastReceivedFrameCount, Frame Check Sequence (FCS) ErrorCount, TransmittedFrameCount, APAverageAccessDelay, AverageAccessDelayBestEffort, AverageAccessDelayBackground, AverageAccessDelayVideo, AverageAccessDelayVoice, StationCount, ChannelUtilisation
Location Configuration Information (LCI)	STAAddress, LatitudeResolution, LatitudeInteger, LatitudeFraction, LongitudeResolution, LongitudeInteger, LongitudeFraction, AltitudeType, AltitudeResolution, AltitudeInteger, AltitudeFraction, Date, AzimuthType, AzimuthResolution, Azimuth
Neighbour Report	BSSID, APReachability, Security, Capability (Cap) CapSpectrumMgmt, CapQoS, Cap Automatic Power Save Delivery (APSD), CapRRM, CapDelayBlockAck, CapImmediateBlockAck, KeyScope, RegulatoryClass, ChannelNumber, PhyType, Neighbor Time Synchronisation Function (TSF) TSFInfo, PilotInterval, PilotMultipleBSSID, RRM EnabledCapabilities

The data gathering with the OMA DiagMon enabler is destined to gather information about 3GPP RATs, but since OMA DM is bearer agnostic, the data gathering can also be realised even if the end-user device is associated with a Wi-Fi AP. The OMA DiagMon data gathering mechanisms can be realised with the OMA DM 2.0 protocol. The OMA DiagMon MO version 1.1 (OMA, 2011b) provides the KPIs shown in Table 15. The green highlighted MOs are useful for the BR.

Table 15: Accessible Key Performance Indicators from OMA DiagMon Managed Object V1.1

MO	Key Performance Indicators
Battery Info	Battery status, battery level (percentage), battery manufacturer, battery version, battery date (date manufactured), battery ID, battery type, battery standby time (estimated standby time in minutes),
Browsing Usage	This function is replaced by the Application Data Usage function of the OMA DiagMon version 1.2.
Data Call and Data Session	Bearer type, data call application, sample duration (in seconds), sample uplink data threshold (minimum number of uplink data (in kilobytes) before data is to be recorded), sample downlink data threshold (minimum number of downlink data (in kilobytes) before data is to be recorded), frequency, report channel, session count, session info (bearer type, data call application and other supported information, e.g. protocol information per session, such as over HTTP, SIP, etc.), session time stamp (start of the session), session duration, uplink data size (in kilobytes), downlink data size (in kilobytes), uplink data speed (in kilobytes per second), downlink data speed (in kilobytes per second), uplink one way delay as defined in (IETF RFC 2679, 1999), uplink one way packet loss as defined in (IETF RFC 2680, 1999), uplink packet delay variation as defined in (IETF RFC 3393, 2002), downlink one way delay as defined in (IETF RFC 2679, 1999), downlink one way packet loss as defined in (IETF RFC 2680, 1999), downlink packet delay variation as defined in (IETF RFC 3393, 2002)

Memory	Short Message Service (SMS) (available memory to store SMSs, in kilobytes), Multimedia Messaging Service (MMS) (available memory to store MMSs, in kilobytes), Random Access Memory (RAM) available (estimation of the currently available RAM, in kilobytes), Ram total (in kilobytes), storageIntAvailable (estimated current available amount of storage space to store data and software, in megabytes), storageIntTotal (in megabytes), storageExtAvailable (estimated current available external amount of storage space, in megabytes), storageExtTotal (External amount of storage space, in megabytes)
Panic Logs	Panic/device crash logs
Restart	Allows the DiagMon Server to remotely restart the device.
RF Metrics	Allows the DiagMon Server to retrieve RF parameters measured by the end-user device. Supported are GSM, UMTS, LTE RF parameters. In this research LTE is the 3GPP RAT that is considered. Therefore, the LTE parameters are listed in the following. Geo location (latitude, longitude, radius), tracking area list, cell list, sampling interval (in seconds, at which sampling interval the diagnostics- and monitoring data are collected at the end-user device), QoS name, QoS lower threshold (data shall be collected when its value is equal or greater than this value), QoS upper threshold (data shall be collected when its value is equal or minor than this value), frequency, report channel, Physical Cell ID (PCID), Downlink evolved Universal Terrestrial Radio Access (E-UTRA) Absolute Radio Frequency Channel Number (EARFCN), Received Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), Received Signal Strength Indication (RSSI), transmission power headroom (device maximum transmit power minus current transmit power)
SMS Options and Usage	Data coding (the coding scheme which is applied to encode the SMS), report status (indicates whether the delivery report is switched on, off, or it is not supported)
MMS Usage	MMS sent (number of MMS sent), MMS received
Near Field Communication (NFC)	Frequency, report channel, NFC RF signal (type and sub-type of technology used, index modulation, communication result (including error code), average bit error and date and time of the failure), NFC software version, failed link attempts, NFC Logical Link Control Protocol (LLCP) failure etc.
User Equipment Setting	Allows the DiagMon server to retrieve UE settings, such as LCD brightness, audio volume etc. The user equipment settings can be updated by the DiagMon server.
Phone Book	Phonebook type (where the type of the phonebook database is stored), phone book used (number of records saved in the database), phonebook free (estimated number of records that could be saved in the database),

The OMA DiagMon version 1.1 has been extended by the OMA DiagMon version 1.2 definition (OMA, 2013b). The additional accessible KPIs are listed in Table 16. Again, the green highlighted MOs are useful for the BR.

Table 16: Accessible Key Performance Indicators from OMA DiagMon Managed Object V1.2

MO	Key Performance Indicators
QoS	The ETSI QoS parameter can be gathered as specified in (ETSI TS 102 250-2, 2011). Examples for ETSI QoS parameters are: Streaming Service Access Time, Streaming Audio Quality, Streaming Video Quality, Streaming Teardown Time, Telephony Speech Quality on Call Basis, Telephony Cut-off Call Ratio, IMS Multimedia Telephony etc.

Sensor	Sensor ID, report condition, frequency, report channel, sensor status, sensor software version, sensor manufacturer, sensor version, sensor manufactured date, sensor geolocation, sensor system location (sensor location within e.g. an industry plant),
Built-in Device Test	Contains the available tests at the end-user device and result codes of the tests.
Device Location	Latitude, longitude, radius, timestamp
Web Browsing Monitoring	Browser name, page Uniform Resource Locator (URL), frequency, report channel, page data (page rendering time)
Application Execution Information	Application info (targeted for application information), open count (count how many times the application is opened), maximum memory usage (specifies if the maximum memory usage of the application has to be recorded), frequency, report channel
Application Data Usage	Allows a DiagMon Server to retrieve specific information about each installed application. It is possible to collect the amount of data and time, the application is used. The availability of this information depends on the support of the application.

The OMA DiagMon version 1.2 extends the DiagMon Framework by adding the trap events framework and by defining the trap events listed in Table 17. The green highlighted traps are useful for the BR. The advantage of using traps, which are available from the OMA DiagMon, is that the data gathering overhead can be minimised. This is because, KPIs are only sent if the observed values show changes that are relevant for the BR, because the BR can register itself on the traps to be notified. Whenever a trap becomes active or inactive, the registered BRs are notified.

Table 17: Traps from OMA DiagMon Managed Object V1.2

Trap	Description
Geographic trap	With this trap a geographic area can be defined. Whenever an end-user device enters this geographic area the trap becomes active and if the end-user device leaves this geographic area the trap becomes inactive.
Received power trap	This trap contains two radio power values expressed in dBm. One radio power value activates the trap if the measured radio power value goes below it and the other radio power value deactivates the trap if the measured radio power value exceeds it. The received power trap is not limited only to 3GPP network interfaces. Non-3GPP network interfaces are supported as well.
Call drop trap	A call drop results in an activation of the trap.
Log capacity full trap	If the defined threshold value (either in kilobytes or percentage of the available log storage) is reached, the trap becomes active.
QoS trap	For all QoS KPIs from Table 16 a lower threshold, which activates the trap if the measured value is equal or greater than the lower threshold, and an upper threshold, which activates the trap if the measured value is equal or minor than the upper threshold.
Hard reboot trap	This trap is activated if a hard reboot is detected. A hard reboot can occur if the power source is abruptly turned off and later on turned on again.

Data speed trap	For downlink and uplink a low average data speed threshold expressed in kilobytes/s, which activates the trap if the measured value is below this lower threshold, and a high average data speed threshold expressed in kilobytes/s, which activates the trap if the calculated value is greater than the high average data speed threshold. The calculated uplink and downlink data speed are average values over a given period of time.

The KPIs are stored in the BR DB and they can be provided to external modules, which are highly reliant on current data, because the accuracy of the underlying data decides on the accuracy of the decisions made by external modules to realise the traffic steering. Through the gathered contexts and policies of the network and the end-user an appropriate decision on traffic steering can be made by the BR in combination with external modules, such as offloading, handover, vertical handover etc.

6.3 Coordination of External Modules

To support well-founded decisions for traffic steering, the BR uses external modules like mobility analysers, vertical and horizontal handover managers, offloading managers etc. The BR provides the architecture and the necessary network elements to integrate these external modules into the decision making process for traffic steering. Through the use of external modules the BR is expandable according to actual and future needs. These external modules can be provided by operators or 3rd parties. The access to the BR DB is provided through UDR FEs used by the BR or through APIs (see chapter 6). Figure 49 shows the information flow between the BR Coordinator and the external modules. The BR Coordinator provides current user context and policies to external modules and triggers the external modules to process the data. The information provided by the BR Coordinator to the external modules enables them to perform a more precise decision making, based on the provided user context and policies.

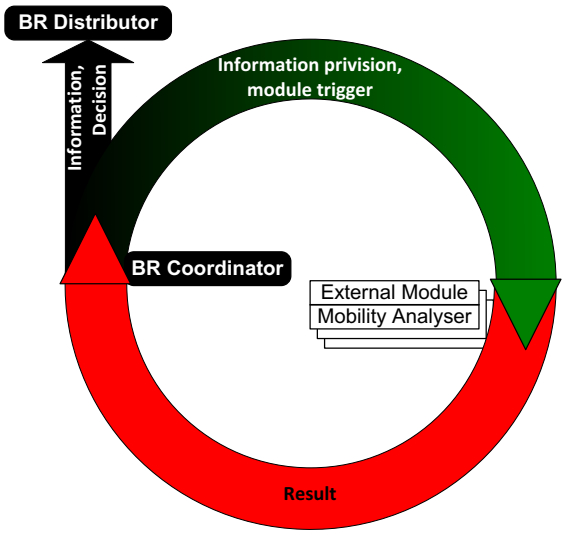


Figure 49: BR Coordinator information flow

The external module returns the decision result to the BR Coordinator. The BR Coordinator may combine the results of multiple external modules to achieve a more precise decision. An illustrative example of this process is given in the following section.

6.3.1 Black Rider Coordinator Example

The coordination function of the BR is demonstrated by considering a mobility analyser and an offloading manager as examples of external modules. First, a mobility analyser is applied to the captured end-user device movement data. The mobility behaviour of the end-user device is analysed and, for example, stable locations, like home and working places, are identified on the basis of the gathered mobility data, provided by the BR DB. The BR provides the relevant data from the BR DB to the mobility analyser. This step is shown in Figure 50 step 1. The mobility analyser is able to make predictions of future movements of the end-user device. The result of the prediction could be expressed for example in the form of a geographic location, where the end-user device is likely to move to, and/or a recommendation of the cell size, depending on the speed of the end-user device. These outcomes of the mobility analyser (Figure 50 step 2) can be used as input data in addition to the up-to-date information provided by the BR DB (Figure 50 step 3) for another external module, which is an offloading manager in this example. The offloading manager uses the provided information from the BR DB (Figure 50 step 4), as well as the results from the mobility analyser as basis for the offloading related decision making process.

From this information, the capabilities of the end-user device could be analysed, for example, which interfaces are supported and which kind of offloading mechanisms are supported, such as LIPA, SIPTO, and IFOM. All the information has to be evaluated, for example, the capabilities of the end-user device have to be justified with the result of the mobility analyser, with the surrounding and available networks in this area and its capacity and utilisation etc. In Figure 50 step 5 the result of the offloading manager is passed to the BR coordinator, which triggers the BR distributor to distribute commands and information to the end-user device.

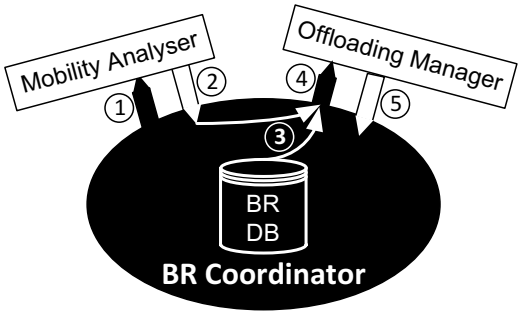


Figure 50: Example of external modules coordination

6.3.2 Influencing Factors for the Black Rider Decision Making Process

Several influencing factors, which could be taken into account for the decision making process, are shown in Figure 51. These examples of influencing factors show their high diversity and the possibilities of gaining benefits in a variety of use cases in different areas.

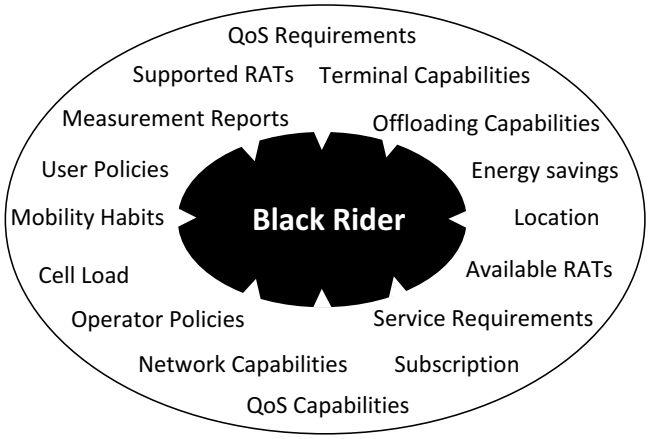


Figure 51: Influencing factors for the Black Rider decision making process

These influencing factors cover a broad spectrum of requirements and demands of the stakeholders. The more factors and parameters are included in the personalised and individual user context, provided by the BR Coordinator to the external modules,

the more precise and accurate are the results of the decision making process. The provided influencing factors have to be evaluated and prioritised by the external modules. The ANDSF, defined by 3GPP, uses XML-based policies (3GPP TS 24.312 V12.3.0, 2013) to describe these influencing factors, parameters, and policies. To be as compatible with existing technologies as possible and to enable a simple and feasible integration of the BR concept into the existing 3GPP architecture, the 3GPP defined XML-based policy description is used by the BR as well. These 3GPP defined XML-based policies can be easily extended with new information types and since the addition of new influencing factors is a strong requirement of the BR concept, the XML-based policies are appropriate for the BR concept.

6.4 Distribution of Commands and Recommendations

The BR supports the distribution of commands and recommendations to the end-user device. Commands are realised through very strict policies that act as commands. Commands contain one or multiple instructions for the end-user device. As a result, the BR commands the end-user device to perform a specific action, such as a handover to a certain RAT or to force the traffic offloading to a particular RAT etc. The distribution of recommendations by the BR includes multiple recommendations. The end-user device can choose one or multiple actions to be performed from the recommendations.

The BR supports the network-controlled, terminal-assisted control mode where the PoD is either located at the terminal or at the network. If the PoD is located at the network, the BR Distributor sends commands to the end-user device. The commands, which are strict policies, have to be followed and finally enforced by the end-user device. If the PoD is located at the terminal, the BR Distributor sends recommendations to the end-user device. These recommendations are also policies, but the end-user device can autonomously select one of the recommendations and enforce the policy.

The support of the two PoDs can be exploited in order to gain benefits for certain kind of use cases. For example, if the end-user device has additional information that enables it to make a more adequate decision than the network is able to do; the PoD can be applied at the terminal. An example of gaining benefit from the control mode, where the PoD is located at the network, is, if the network is highly utilised and

therefore end-user devices have to be forced to offload traffic through handover or traffic offloading procedures to other RANs. In such a case, it is very important that the BR is able to command the end-user devices which execute the commands promptly. To transfer the control messages – the commands and the recommendations – between the BR Distributor and the end-user device, the OMA DM version 2.0 (OMA, 2013a) is used in this research. As defined in 4.10, the Hotspot 2.0 Release 2 protocol is considered as the secondary traffic steering protocol. The fact that this protocol is able to send commands and recommendations to the end-user device is considered, but no statements can be made yet about the real-time capability of this protocol.

6.4.1 OMA DM Session Establishment

It is essential for the traffic steering that the commands and recommendations can be sent to the UE in a prompt and real-time manner. If the commands and recommendations arrive at the end-user device too late, they are either useless or the effect of the commands and recommendations may be marginal or even disadvantageous. The commands and recommendations are sent via the OMA DM version 2.0 protocol. Therefore, the BR Distributor has access to the OMA DM server to enable the OMA DM version 2.0 protocol at the BR side. The problem that may affect the real-time behaviour is that only the client can establish an OMA DM session. Although the BR Distributor can trigger the end-user device via the OMA DM server to establish an OMA DM session by sending a DM Notification, the client always initiates the DM session as shown in Figure 52.

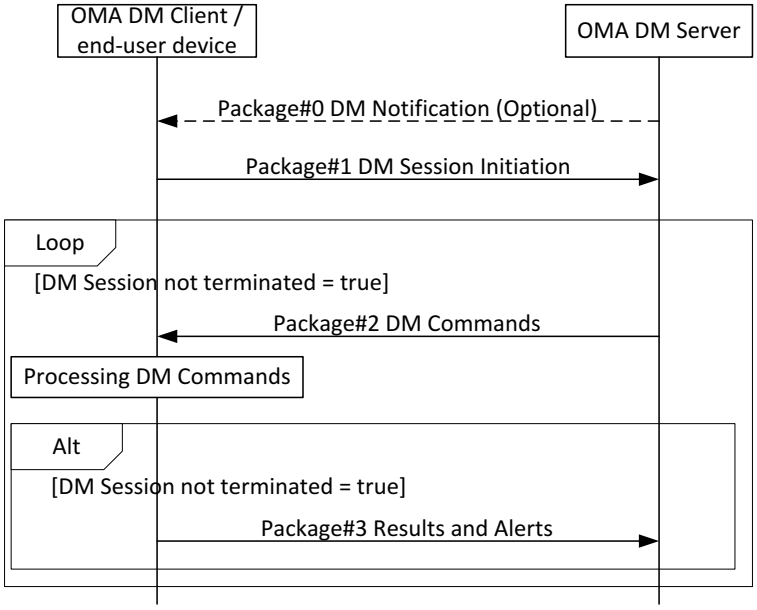


Figure 52: DM session establishment and package flow

Although end-user devices cannot continuously listen for connection from the OMA DM server or are not willing to open a port, unsolicited messages (notifications) can still be received by most of the end-user devices. One way of sending notification by the OMA DM server is to send an SMS message to trigger the end-user device to initiate a DM session. Table 18 shows the possible transport mechanisms for DM notifications defined in (OMA, 2013a). The most common way of sending a notification to the end-user device is the use of connectionless WAP Push sent via an SMS. Another approach for sending notifications is the Google Cloud Messaging (GCM) for Android. The GCM is either sent over Extensible Messaging and Presence Protocol (XMPP) or HTTP. But this solution is limited only for end-user devices running Android as the Operation System (OS). Other transport mechanisms are allowed but not explicitly indicated in (OMA, 2013a).

Table 18: Variants of transport mechanisms for DM notifications

Transport Mechanisms	Protocol / Method	Description
Connectionless WAP Push	SMS or UDP.	If WAP Push sent over UDP the end-user device has to listen to the defined port number 2948.
Google Cloud Messaging (GCM) for Android	XMPP or HTTP	There is a GCM entity at the end-user device and at the network side. After a registration process of the end-user device the OMA DM server can send notifications to the GCM network entity, which sends the notification to the GCM entity at the end-user device where the notification is passed to the OMA DM Client.

Other transport mechanisms	Are allowed	

The DM commands are processed by the end-user device in a sequential order. The supported commands by the OMA DM version 2.0 (OMA, 2013a) are listed in Table 19.

Table 19: Supported DM commands by the OMA DM version 2.0

Command	Description
HGET	If the OMA DM client receives this command it requests specific MO(s) from the OMA DM server with a HTTP GET request.
HPUT	If the OMA DM client receives this command it sends a HTTP PUT request sending the requested data.
HPOST	If the OMA DM client receives this command it sends a HTTP POST request sending the requested data.
DELETE	If the OMA DM client receives this command it will delete the specified node in the tree including all child nodes.
EXEC	If the OMA DM client receives this command it will execute the executable node specified in this command.
GET	If the OMA DM client receives this command it sends a HTTP POST request sending the requested MO. The MO is not included in package#3, instead it is included in the HTTP message body, e.g. the HTTP POST message.
SHOW	If the OMA DM client receives this command it initiates a User Input (UI) session between the web browser and the web server.
CONT	If the OMA DM client receives this command it continues the DM session with the specified OMA DM server.
END	If the OMA DM client receives this command it terminates the DM session. Only the OMA DM server can terminate a DM session. The OMA DM client is not allowed to terminate a DM session.
DEFAULT	If the OMA DM client receives this command it can use the specific address included in the DEFAULT command to capture configuration if the configuration is missing in the device.
SUB	If the OMA DM client receives this command it reports any changes in the specified DM tree part. Note that this subscription causes the OMA DM client to report any changes. This contrasts with the DiagMon traps, where thresholds can be defined and only if the values exceed or fall below the defined thresholds, the changes are reported.
UNSUB	If the OMA DM client receives this command it does not anymore report any changes in the DM tree part to the OMA DM server.

6.5 Real-time Capability in Time-Critical Situations

Traffic steering often faces time-critical situations where the commands have to be sent from the BR to the end-user device in a prompt manner. Time-critical situations are for example handover and offloading procedures. In section 6.4.1 the problem that can affect the real-time behaviour has been mentioned. Actually, only the client can establish an OMA DM session. The BR is able to force the OMA DM client to

initiate and establish a DM session, but the BR is not able to immediately send commands to the end-user device. This problem, affecting the real-time behaviour, is tackled by the trap framework provided by the OMA DiagMon version 1.2. Thresholds can be defined in the traps. If the values exceed or fall below the thresholds, the traps are set active or inactive and the subscriber of the trap gets notified. The subscriber in this research is the BR. The notification sent to the BR is sent over a DM session. Therefore, whenever a notification has to be sent to the BR, the OMA DM client has to either initiate a DM session in order to send the notification to the BR or send it over an already established DM session. The occurrence of a time-critical situation is accompanied by exceeding or falling of values below the defined threshold. After the value reporting to the BR, the traffic steering decision making process is started by the BR. While the traffic steering decision making process takes place, the DM session still remains. The DM session is not terminated by the BR. If the commands or recommendations are ready to be sent to the end-user device the still established DM session can be used, which is real-time capable once initiated. Even if data is gathered through the IEEE 802.11k or through the UDR/BR DB-based concept, indicating the impending of a time critical situation, the BR is able to force the OMA client to initiate a DM session with the package#0, the DM Notification message, before triggering the traffic steering decision making process. As a result, also time-critical situations can be handled by the BR in real-time so that the end-user device receives the commands and recommendations in time to execute the command or the selected recommendation.

6.6 Black Rider at the Cloud

One instance of a BR at the cloud is serving multiple end-user devices. Every end-user device is steered individually through the mobile HetNets by a BR instance. The reason to choose a one-to-many relationship between the BR and the end-user devices is because of scalability issues. A one-to-one relationship between a BR and an end-user device would not scale with millions of end-user devices. The BR is located on a virtual machine at the cloud. But since the end-user device is mobile this ability of mobility can cause increased delay and latency, because the movement of the end-user device can cause an increase of the resulting geographic distance between the end-user device and the serving BR instance at the cloud. To prevent a

high delay, due to long geographic distances between the serving BR and the end-user device, an instance of a BR has its defined serving area, such as a SGW or an MME has its serving area. If an end-user device is exiting the serving area of its actual BR instance and enters the serving area of another BR, the contexts and policies have to be handed over from the serving BR towards the target BR. As a result, the delay can be reduced and the target BR is able to serve the end-user device without building up all the contexts and policies from scratch.

6.7 Summary

The architecture of the BR with its building blocks and necessary concepts to provide smart, intelligent, and individualised traffic steering have been described and explained in this chapter. The description of the building blocks provides detailed insights into the necessary and involved network elements. Data gathering is a central and very basic necessity in order to provide the BR services based on this information. Therefore, the data is first classified into different classes containing context and policies of the network and the end-user device. As a second step, the various data gathering variants are presented. Because a HetNet environment is supported by the BR, multiple variants of data gathering have to be supported. KPIs have been described and information is provided which KPIs are available with which data gathering variant. The BR provides the support of the integration of external modules that analyse specific parameters and provide results back to the BR. These external modules are triggered and coordinated by the BR which is shown by an example. A selection of influencing factors and parameters that influence the decision making process of the BR is provided. To provide the commands and recommendations towards the end-user device, protocols are necessary to enable the BR to distribute these commands and recommendations. The considered protocols are the OMA DM version 2.0 and the Hotspot 2.0 Release 2. The BR is real-time capable with the OMA DM version 2.0. The standardisation of the Hotspot 2.0 Release 2 is not yet finished and no statements can be made yet about the real-time capability of this protocol. Therefore, Hotspot 2.0 Release 2 is considered as the secondary protocol. The BR is located at the cloud. One BR instance serves multiple end-user devices and has a defined serving area. To prevent high delays from the BR to the end-user device, in order that the end-user device moves away from the BR location,

the contexts and policies of the specific end-user device is handed over to the target BR in the new serving area.

7 Black Rider Use Cases

In the following, the BR architecture is applied to several use cases to show how the BR concept can be systematically exploited to gain benefits. The handover and offload use case shows two control variants. One is purely network-controlled, terminal assisted where the PoD is also located at the network and the other is network-controlled, terminal assisted where the PoD is located at the terminal. The energy efficiency use case demonstrates that a user selected policy can impact the handover behaviour and access network selection resulting in a decrease of energy consumption. The information provision to 3rd party use case shows by example how the provisioning of information to 3rd parties can improve the accuracy of their offered service. The elimination of the need of adding an offset to the SSDL in a HetNet environment with the application of the BR, to prevent the macro cells from an over-utilisation and the small cells from an under-utilisation, is shown in 7.4.

7.1 Handover and Offload Use Case

The handover and offload procedure are similar to one another. An offload and handover procedure, no matter if it is a horizontal handover or a vertical handover, has to progress three phases:

System discovery: The end-user device sends the end-user device context to the network. Usually this is done as soon as changes occur to the context. The end-user device retrieves general system information from the surrounding access networks over the broadcast channel. The end-user device performs measurements on several parameters, depending on the RAN connected to, and sends measurement reports to the network. Figure 53 illustrates the distribution of the policies and context information of the end-user device- and network-side. In a pre-handover or offload situation, the end-user device sends its end-user device context to the BR by using one or multiple data gathering variants, which have been previously discussed in section 6.2. The general system information is sent frequently by the surrounding cells to the end-user devices within reach over the RAT specific system information broadcast channel.

Handover and offload decision: There are two variants dependent on where the handover or offload decision is made. The first variant uses the network-controlled, terminal assisted mode with the PoD located at the network. The BR performs the handover or offload decision based on the contexts and policies. The individual handover or offload command is sent to the end-user device. The second variant is where the network-controlled mode is terminal assisted and the PoD is located in the terminal. As in the first variant, the BR performs the handover or offload decision based on the contexts and policies, but instead of one handover or offload decision the result may contain multiple handover or offload recommendations which are sent as individualised recommendations to the end-user device.

Handover or offload execution: The handover or offload decision is enforced with a handover or offload execution. The BR can either send individual control recommendations or individual control commands to the end-user device. If control recommendations are sent from the BR to the end-user device, then the PoD is located at the end-user device which is then in charge of making the final decision. Therefore, the end-user device considers the end-user device policies and context information. If a control command is sent from the BR to the end-user device the PoD remains at the network side and the BR is in charge of the decision making process. With this concept the PoD can either be located at the end-user device or at the BR. As a result of supporting both approaches, the best decision can be made either at the network- or end-user device-side.

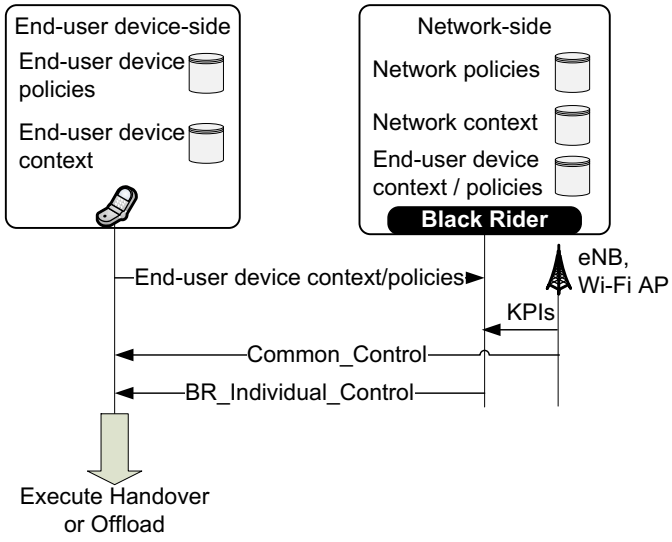


Figure 53: Usage of contexts and policies before and during a handover or offload situation

Table 20 shows a comparison of the handover and offload procedures with and without the application of the BR. Through the application of the BR it is possible to locate the PoD at the network, the individual and personalised control is enabled and commands can be sent to the end-user device to enable to command the end-user device. As a result of applying the BR, an individual and personalised traffic steering for handover and offloading is possible.

Table 20: Handover and offload comparison with and without the application of the BR

	BR applied	BR not applied
PoD at the network	X	
PoD at the terminal	X	X
Individual Control	X	
Command the terminal	X	
Recommendations	X	X

7.2 Energy Efficiency Use Case

It is essential that the energy consumption of an end-user device is minimised to extend the duration of a battery charge. For the energy efficiency use case, the operator provides different pre-defined energy models to the users, from where they can choose from. Example definitions of such pre-defined energy models are given in Table 21.

Table 21: Energy model definition

Energy Model	Classification	Used RATs
LEC	Low Energy Consumption (LEC) Rather low QoS	3G, Wi-Fi
MEC	Medium Energy Consumption (MEC) Medium QoS	LTE, Wi-Fi
HEC	High Energy Consumption (HEC) High QoS	LTE
AUT	Depending Due to mobility behaviour analysis of the end-user device and the history of selected energy models by the user, the BR can perform an automated (AUT) energy model selection which results in an appropriate RAT selection.	

All the user selectable models and policies may have a lower priority than the operator defined policies. To compare these energy models, an area with different RATs and their coverage ranges is defined. This area is applied for all energy models.

To be able to compare the energy consumption of the different user selected energy models, the end-user devices are moving straight from one end of the area to the other end, as it is depicted in Figure 54.

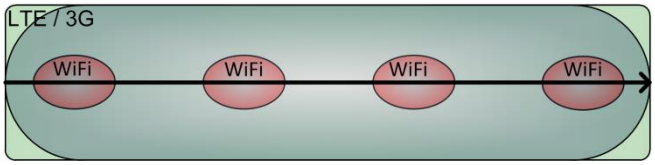


Figure 54: Coverage area for the energy efficiency comparison

The provided RATs are LTE, 3G, and Wi-Fi and the network operator is the same for the three network types. The coverage is depicted in Figure 54 and shows that LTE and 3G are covering the whole travelling path of the terminal and Wi-Fi is deployed in hotspots. It is assumed that the travelling lasts one hour and that it is possible to access 4 times through Wi-Fi for 4 minutes. In (Huang et al., 2012) a data transfer power model for the three RATs – LTE, 3G, and Wi-Fi – has been derived as shown in Table 22.

Table 22: Data transfer power model (Huang et al., 2012)

	α_u(mW/Mbps)	α_d(mW/Mbps)	β(mW)	Idle(mW)
LTE	438.39	51.97	1288.04	594.3
3G	868.98	122.12	817.88	374.2
Wi-Fi	283.17	137.01	132.86	77.2

β is the base power when throughput is 0. It has been observed that a linear model fits well for both uplink and downlink. Uplink throughput is t_u [Mbps] and downlink throughput is t_d [Mbps]. The uplink and downlink power consumption per Mbps α_u and α_d are given in Table 22. The power level [mW] for the uplink is shown in equation (14) and the power level for the uplink is shown in equation (15).

$$P_u=\alpha_u t_u+\beta \quad (14)$$

$$P_d=\alpha_d t_d+\beta \quad (15)$$

The BR concept is applied using the three energy models LEC, MEC, and HEC from Table 21. The results are compared with and without the use of the BR. It is assumed that three users travel the path, shown in Figure 54, which lasts one hour. Each of the users has selected a different energy model and adapts the communication activities

according to the selected energy model. The activities of the three energy models are listed in Table 23. In case the BR concept is applied, Wi-Fi can be used as alternative radio interface, whenever available. If the BR is not used, Wi-Fi is not used as an alternative RAT. The reason for this is that today many users disable Wi-Fi at the end-user device, because of a bad QoS experience when switching to Wi-Fi. This is also caused due to the behaviour of the end-user device. For example, the iPhone from Apple is switching to Wi-Fi whenever a Wi-Fi AP is detected, no matter how bad the QoS is. Samsung end-user devices have a better behaviour. If a Samsung end-user device detects a Wi-Fi AP, the connection is tested with a Google server. Only if the delay and throughput reach certain thresholds, Wi-Fi is used for the connection. However, many users disable the Wi-Fi at the end-user device to prevent them from a bad QoS and to avoid the ping-pong effect of switching between Wi-Fi and 3G/LTE. The activities are described either by duration (s), if the activity is continuous, or by the number of events, if the activity is of a non-continuous kind, such as web browsing and sending and receiving emails. The last two columns show the used RATs per activity with and without the usage of the BR and the idle time in [s] of the RAT.

Table 23: Activities per energy model

HEC		**With BR**	**Without BR**
Activity	**Duration [s]**	**RAT / Duration [s]**	
Skype	720	LTE	LTE
YouTube	450	LTE	LTE
Audio streaming	1200	LTE	LTE
Idle LTE		1230	1230
MEC		**With BR**	**Without BR**
Activity	**Duration [s] / Number**	**RAT / Duration [s]**	
Skype	600s	LTE	LTE
Web browsing	14	LTE	LTE
Web browsing	20	Wi-Fi	LTE
Tx text email	2	Wi-Fi	LTE
Tx email with attachment	3	Wi-Fi	LTE
Rx text email	5	Wi-Fi	LTE
Rx email with attachment	3	Wi-Fi	LTE
Idle Wi-Fi		835.5	
Idle LTE		2031.3	2953.3
LEC		**With BR**	**Without BR**

Activity	Number	RAT / Duration[s]	
Web browsing	2	3G	3G
Web browsing	10	Wi-Fi	3G
Tx text email	3	Wi-Fi	3G
Tx email with attachment	1	Wi-Fi	3G
Rx text email	5	Wi-Fi	3G
Rx email with attachment	3	Wi-Fi	3G
Idle Wi-Fi		918.3	
Idle 3G		1629.3	3488.2

For Skype, YouTube, and streaming audio, a continuous traffic stream of 64Kbps, 770 Kbps according to (Zink et al., 2008) and 128Kbps is assumed. Furthermore it is assumed that on average website request has a size of 1KB, an average website has a size of 1.5 MB (Souders, 2014), a plain text email has a size of 0.2MB and an email with an attachment has a size of 4MB. The necessary power for an activity for the uplink and downlink P_A [W] is calculated using equation (16), following from (14) and (15).

$$P_A = \frac{\alpha_u t_u + \alpha_d t_d + \beta}{1000} \quad (16)$$

If an activity consists of a number of executions, such as web browsing and emails, the number of execution is multiplied by the P_A term. The interface sends and receives data in a sequential order. E_{tot} [J] is the total energy consumption of all the activities as given in (17).

$$E_{tot} = \sum_{i=1}^{n} P_{Ai} * \Delta t_i \quad (17)$$

Table 24 shows the uplink and downlink throughput in Mbps for the LTE, the 3G technology, which includes mainly HSPA technology, and the Wi-Fi access technology. The LTE and 3G throughput values are from (RootMetrics, 2014). RootMetrics made over 4 Million tests throughout the USA to get this throughput data. These throughput values are from the 1st half of 2013. The Wi-Fi values have been measured using an IEEE 802.11n standard enabled Wi-Fi Network Interface Cards (NICs) supporting 2x2 MIMO at each of the 5 clients and also an IEEE 802.11n enabled AP supporting 2x2 MIMO. An AP was at any time serving 5 STAs. The measured results are similar to the Aruba measurements in (Aruba, 2008) stating an approximately average throughput

of 100Mbps over the tested APs with one STA. The Wi-Fi throughput varies because of the used MIMO, the number of spatial streams, the number of STAs etc. IEEE 802.11ac AP and STAs are nowadays still very rare; therefore, the comparison is done with the IEEE 802.11n standard.

Table 24: Uplink and downlink throughput per Radio Access Technology

RAT	Throughput [Mbps]	
	Downlink	Uplink
LTE	16.45	8.75
3G	4.3	1.1
Wi-Fi	22	22

Figure 55 shows the results of the energy consumption per energy model, with and without applying the BR concept. Since the energy model HEC is not supposed to save any energy, the required power is the same for both solutions - with and without applying the BR. For the other two energy models, the MEC and the LEC, the solution with the BR saves energy.

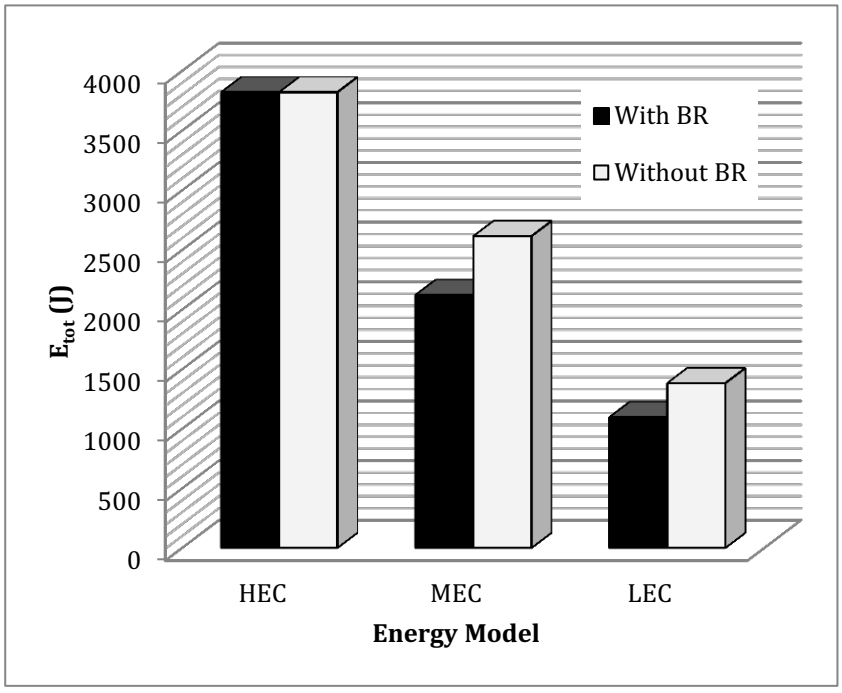


Figure 55: Energy consumption per energy model with and without application of the Black Rider

Table 25 shows the results of the energy consumption calculations per energy model when the BR is applied and if it is not applied. With the application of the BR the power consumption is 18.85 percent lower for the MEC and 20.66 percent lower for the LEC energy model. This is due to the use of Wi-Fi when the BR is applied. The

detailed calculations of the energy consumption per energy model with and without the application of the BR are given in the Appendix A

Table 25: Energy consumption per energy model with percentage improvement values

Energy consumption [J]	HEC	MEC	LEC
With BR	3820.93	2122.20	1096.87
Without BR	3820.93	2615.16	1382.59
Improvement in Percentage	**0**	**18.85**	**20.66**

The required energy consumption to perform a handover or offload to a different RAT is not considered in this evaluation. Since the BR can provide the appropriate information about frequencies and IDs of the actually available RATs to the end-user device, there is no need for the terminal to perform a long scan of the surrounding networks. The authentication can be shortened, because the required information is provided by the BR and the use of Hotspot 2.0 at the Wi-Fi RAT. With these improvements provided by the BR, the necessary power to perform a vertical handover or offload between LTE/3G and Wi-Fi can be minimised compared to the case where the BR is not applied.

7.3 Information Provision to 3rd Party

Geolocation information is used for several services, such as CDNs, cloud balancing, direct marketing, context-sensitive content delivery etc. To provide geolocation information the widely used IP geolocation approach is applied, where mappings of IP address blocks to geolocations are stored in databases. Another technique is the active way of getting geolocation information, by measuring the delay. With this technique, more accurate geolocation information can be achieved, but with the lack of scalability, high measurement overhead, and very high response time ranging from tens of seconds to several minutes to localise a single IP address (Poese et al., 2011).

The BR can store very precise location information of the attached end-user device on an EPS in its serving area. For 3GPP technologies, the end-user device position is known at cell identity level, if a bearer is active. When an end-user device communicates via a WiMAX access network, the end-user device is known at Base Station (BS) identity level. If Wi-Fi is chosen for the access, the end-user device location is known, for example, on cell level using the BSSID. Another possibility is that the end-user device is aware of its location, for example through GPS, and sends

the co-ordinates over the OMA DM protocol to the BR. With this information the BR knows the actual location of stationary and moving end-user devices even if the IP address remains the same. Whenever the location related data sets in the BR DB are created or updated the geolocation information is used to create or update the corresponding country, city and zip code entries. This information can be provided by the BR DB to 3rd parties, such as CDN companies, to improve their service accuracy. Figure 56 illustrates the two variants how 3rd parties are able to get access to the BR data sets. One possibility for 3rd parties to access the data sets is over the nwIBR DB interface.

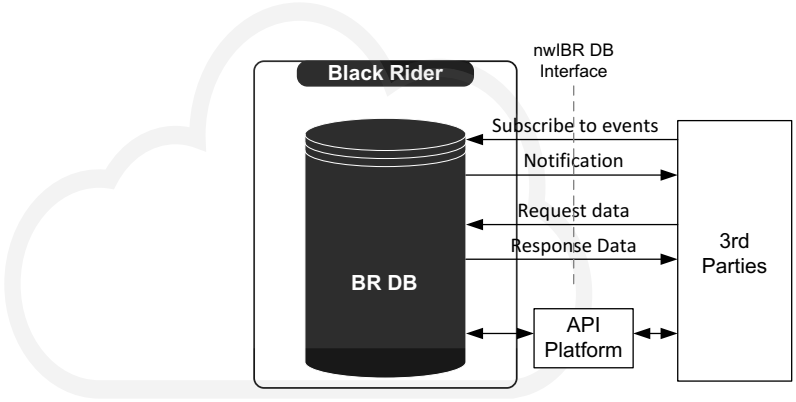


Figure 56: Information provision to 3rd parties

To keep the 3rd party database updated either a data request is sent to the BR DB which responds by sending the requested data or the 3rd party can subscribe to events, such as the change of the country, city or zip code data sets and get notified by the BR DB in case of event occurrence. The BR DB has to provide an appropriate FE to communicate with the 3rd party via the nwIBR DB interface. The other possibility for 3rd parties to get access to data sets of the BR is with the use of an API provided by an API platform. By providing this geolocation information to 3rd parties enables them to improve the accuracy of their offered services.

7.4 Macro and Small Cell Interference Offset Improvements

In section 3.3.2 the problem of cell selection within HetNets is described. To select a cell, the SSDL of different cells is measured by the end-user device and the cell with the strongest signal strength is selected. In most cases this selection mechanism results in selecting the macro cell, which leads to an over-utilisation of the macro cells and to an under-utilisation of small cells, because the small cells are rarely selected.

To solve this problem, the CRE is introduced which adds an offset to the SSDL of small cells with the result that end-user devices increasingly select the small cell, even at the edge of small cells. This mechanism depends upon the added offset.

With the application of the BR this cell selection problem could be improved by additionally considering the location of the end-user device. Since the locations of the small cells are known, this information can be used to define a geographic area, where it is appropriate for an end-user device to select the small cell. This would decouple the cell selection from the SSDL value and would increase the accuracy of cell selection. With the BR another parameter could be taken into account: the location of successfully received data from the small cell. If the location of successfully received data is gathered, areas within the reach of the small cell could be identified where receiving data is possible, as well as areas where the receiving of data is not possible anymore. Furthermore, the speed of the end-user device can also be considered to prevent the end-user device from selecting a small cell if the speed is too high.

7.5 Summary

The BR use cases give an impression of the broad area of application, where the BR can be applied and the BR concept can be systematically exploited to gain benefits. The more parameters are considered and processed by the BR in its external modules, the more possibilities are created to support the BR concept using the individual context of end-users. As a result, new services can be defined to improve the experience for the mobile end-user due to an improved traffic steering concept. Due to the application of the network-controlled mode, the terminal-assisted support, and the flexible PoD either at the network or at the terminal, the BR can be applied in a variety of use cases. The benefits, of applying the BR for handover and offload decisions, have been outlined in this chapter. The achieved improvements to energy savings at the end-user devices through the application of the BR have been clearly shown, as well as the benefits the BR could bring to 3rd parties with the example of providing more accurate geolocation information. Furthermore, the added offset to the SSDL can be eliminated by the BR through the consideration of the end-user device location parameters. The selection of small cells and macro cells within HetNet environments can be even more improved when considering further parameters such

as the end-user device speed and the combination of the location and received data at the end-user device.

8 Simulation

For the simulation the offload use case 7.1 is selected to be implemented. The offload use case was selected because it has a big influence on the traffic steering. The offload use case has the potential to prevent RATs from overutilisation and to distribute the traffic load over multiple RATs. This is a very important capability considering the predicted huge increase of traffic in mobile networks.

The aim of the simulation is to clarify the following questions:

- Study of the logical and structural behaviour. This contains the investigation of the BR architecture. Is the BR architecture coherently executable and implementable?
- Performance study. Is the BR able to improve the KPIs, such as the number of offloads, the throughput, the received KB? The investigated KPIs are defined in 8.2.5.
- Does the application of the BR improve the confidence interval?

The parameters that the BR takes into account are reduced to the ones with the most assessable impact. This is due to the time and resource limitations within this research. These parameters are the location of the end-user device, the history of the end-user device's movement, the packet receiving correlated to the location and the throughput.

The network simulator ns-3 has been selected for the simulation very early in the research. The comparison of open source simulators is given in Table 26. Commercial or academic network simulators that are not freely available, such as Optimized Network Engineering Tool (OPNET) are not considered, because of the reduced financial possibilities within this research.

Table 26: Simulator comparison

Criteria / Simulator	Ns-2	Ns-3	OMNET++
LTE Module	Since 2010	Since 2011	Since 2014
EPS	Not fully	Fully	Not fully
Wi-Fi Module	X	X	X

An LTE module (eNB, UE) and an EPS module (MME, SGW, PDN GW) were already implemented in ns-3 in 2011. The implementation was done by the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) in co-operation with Ubiquisys. In the following years this LTE/EPS module was continuously developed further through Google Summer of Code (GSoC) projects under the lead of the CTTC and through the CTTC itself. It is a very stable and continuous development until today. The development of the ns-2 and OMNET++ LTE and EPS modules are less rapid in progress than the one in ns-3. Since today ns-3 has the most comprehensive LTE and EPS simulation environment implemented. The ns-3 is an open source, discrete-event network simulator for internet systems, targeted primarily for research and educational use. The ns-3 is free software, licensed under the GNU GPLv2 license, and is publicly available. The ns-3 provides multiple methods to verify and validate the implemented simulation modules such as tests and traces.

- Build Verification Tests are aimed to ensure that the build is working.
- Unit Tests are detailed tests to ensure the accuracy of a piece of source code in isolation.
- System Tests involve multiple simulation modules to ensure the interworking.
- Examples are simulation setups that usually generate traces to ensure the accuracy of a simulation module interacting with other modules. The traces can be used to verify, if the simulation model is valid and congruent with the real system.
- Performance Tests are used to check if a particular part of the simulation system completes in a reasonable time.

The ns-3 offers multiple tracing variants:

- NS_LOG is a logging method that can be used to write information into a file. This information can be evaluated afterwards through scripts that use for example the Linux commands grep, sed or awk to parse the information.
- Ascii files can be generated. These files contain similar information as Packet Capture (pcap) files, but in ASCII format.
- pcap files can be generated which can be analysed afterwards with a network traffic analyser software such as wireshark.

To simulate the BR the main ns-3 modules are the Wi-Fi and LTE simulation modules. These two simulation modules have dependencies on other simulation modules. The LTE module is documented in (Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), 2014) and the Wi-Fi module is documented in (ns-3 community, 2014a). The documentations of all the ns-3 modules are available from (ns-3 community, 2014b). The validation and verification of the LTE and the Wi-Fi modules are given with the unit tests, the system tests, the examples that are provided and documented within the modules. Furthermore, the LTE module is validated in (Baldo et al., 2011) and the Wi-Fi module is validated in (Baldo et al., 2010). The BR is a new module which has been implemented as part of the thesis in the ns-3 simulator enforcing the policies and the behaviour described in section 8.2.3 and 8.2.4. The Non-Access Stratum (NAS) EPS Mobility Management (EMM) and EPS Session Management (ESM) protocols between the UE, the eNB and the MME to establish default and dedicated bearers have been implemented within this research as well. The mobility protocol of the EPS and for Wi-Fi access networks and EPC interconnection is the GTP, which had been implemented and integrated into the ns-3 during this research work (Frei et al., 2012b) (Frei et al., 2011c).

To calculate the confidence intervals and the average values of the simulation runs, the freely available, high level interpret language GNU Octave is used (GNU Octave, 2014). Several GNU Octave scripts have been written during the research work. To generate the diagrams, the freely available, portable command-line graphing utility Gnuplot is used (Gnuplot, 2014). To automate all these calculations and for generating diagrams, several bash scripts have been written.

8.1 Traffic Steering Scenarios

In the following the three defined traffic steering scenarios are discussed which are used to verify the advantages of the BR. They consist of one traffic steering variant considering the situation nowadays, and two traffic steering variants where the novel BR is applied.

8.1.1 Variant 0: How Offloading and Inter-System Handover Work Nowadays

This variant uses the terminal-control mode, with the PoD located at the terminal. Today, most cellular phones support multiple interfaces like 3GPP interfaces such as

2G/3G/4G and non-3GPP interfaces, where Wi-Fi is the most common one besides WiMAX. Since in this thesis Wi-Fi is used as the non-3GPP technology, WiMAX is not considered. Today, it can be a struggle to connect to a Wi-Fi hotspot with end-user devices. Subscribers who want to connect to a Wi-Fi access network might go through several manual steps, such as searching or scanning for access networks, selecting an access network, and enter credentials. Usually the credentials are entered by using a web browser. A very common method to authenticate subscribers at a Wi-Fi hotspot, even in the roaming case is the use of a captiva portal. Every packet sent from an unauthorised device is intercepted until the user opens a web browser and tries to access the internet. Whichever Uniform Resource Locator (URL) is entered by a user in a web browser, the user is redirected to a website where payment is required, or the user has to enter the credentials manually, or simply an acceptable use policy is displayed. Captiva portals can be set up using the Extensible Markup Language (XML) -based authentication protocol Wireless Internet Service Provider roaming (WISPr) and Universal Access Method (UAM) in combination with a Remote Authentication Dial-In User Service (RADIUS) server responsible for the AAA. With WISPr and UAM it is possible to automate the login process. The automation of the login process with WISPr is supported by the mobile i Operation System (iOS) from Apple and by many third party applications called Connection Managers (CMs) to provide this capability of automatically login to a Wi-Fi hotspot. These solutions are not available to all devices and are offered by a limited range of carriers. Therefore, these automated solutions are far from widespread.

Without having automated processes to connect to Wi-Fi access networks or change the Wi-Fi access point during roaming, a seamless offload and handover solution is not possible. Only non-seamless offload and handover is then possible. The present solutions also lack an automated access network selection function which is eminent for a seamless offload and handover solution.

8.1.2 Variant 1: BR@Cloud with Point of Decision at the Terminal

This variant is depicted in Figure 57. The network-controlled, terminal assisted mode is used with the PoD located at the terminal. This variant is a mixed control mode, where the network has the control to narrow the selection, but the terminal owns the control of making the final decision. The S14 interface and the Hotspot 2.0 Release 2

solutions assist the end-user devices with the access network selection by providing information and policies about the available surrounding access networks. The end-user device receives the information and policies from the BR and has to process this data to finally make a decision. The decision making process contains not only the received information and policies; it includes also the user preferences stored in the end-user device itself. Especially in time critical situations such as handover and offload processes the decision should be made as quick as possible.

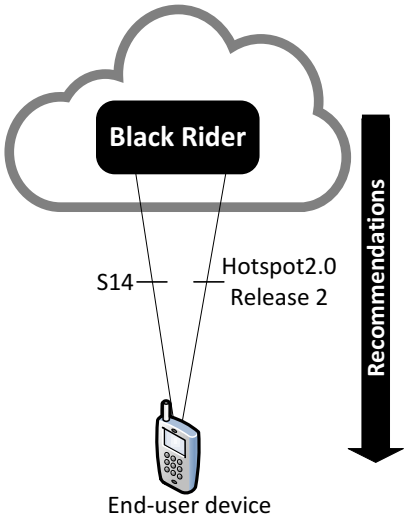


Figure 57: Variant 1: Network-controlled and terminal assisted mode, PoD located at the terminal

8.1.3 Variant 2: BR@Cloud with Point of Decision at the Network

This variant is a network-controlled, terminal assisted mode where the PoD is located in the network. Therefore, the network owns the complete control and the terminal has to execute the commands from the network. Variant 1 uses the S14 and the Hotspot 2.0 Release 2 functions only for providing assistance to the end-user device with the access network selection by providing information and policies about the surrounding, available access networks to the end-user devices. Variant 2 differs from variant 1 in the kind of policies sent to the end-user devices. The policies of variant 1 can narrow the selection of the access network, but there is no way of commanding the UE to execute commands promptly. Variant 2 uses the S14 and the Hotspot 2.0 Release 2 functions to provide commands and recommendations to the end-user devices, as can be seen in Figure 58. These strict policies are then commands provided by the BR to the end-user devices where these commands are promptly executed. The interfaces to provide the commands and recommendations are the

same as with variant 1. The end-user device is not in charge of making the decision; the decision has already been made by the BR. The BR has also access to the user preferences and therefore, compared to variant 1, there is no drawback or lack of information in the decision making process by the BR. Commands have the advantage that they can be executed very quickly, because there is no need to process any data or perform a decision making process. This is important in time critical situations, such as handover or also traffic offloading processes. Therefore, the end-user device does not lose any time, as it would in variant 1.

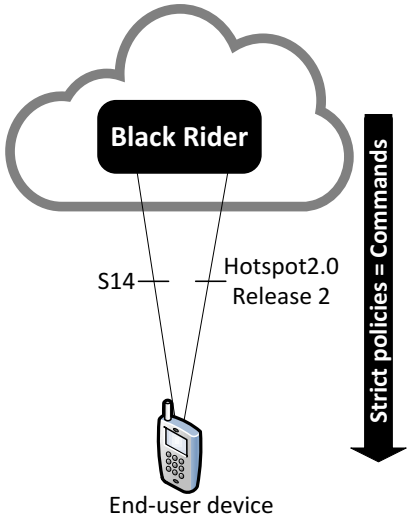


Figure 58: Variant 2: Network-controlled and terminal assisted mode, PoD located at the network

8.2 Simulation Setup

The network architecture for the simulation is shown in Figure 59. The PDN GW and the SGW are collocated in one network entity. For the connection of the Wi-Fi APs a so called Access Point Gateway (APGW) has been implemented during this research in the ns-3 and integrated into the PDN GW / SGW device. The Wi-Fi APs are connected to the APGW over the S2a interface where GTP is the deployed mobility protocol. The eNBs are connected to the PDN GW / SGW over the S1-U interface using the GTP protocol. The S1-U interface and the S2a interface have a delay of 7 ms. The connection link between the remote host and the PDN GW / SGW has a delay of 20 ms. The control plane path with the S1-MME and the S11 interface have both a delay of 2.5 ms. The delay between the STA and the AP is 4 ms and the delay between the UE and the eNB is 3 ms. The BR has a delay of 7 ms (UE) and 8 ms (STA) towards the end-user devices which includes the delay between the eNB and the UE and the AP

and the STA. These 3GPP related delay values are taken from operator measurements. The Wi-Fi delays have been measured in a laboratory setup with different distances between the AP and the STAs. From the results, the average has been calculated. Every end-user device is a multi modal device containing an LTE and a Wi-Fi interface. Therefore, the end-user devices are both – UEs and STAs. During the simulations, the network interfaces can be actively used in parallel. But one bearer can only be transmitted either on the LTE or Wi-Fi interface. Therefore, one bearer is never distributed over the 2 different RATs. The BR has connections to the eNBs, the APs, and the end-user devices. The remote host is used to send/receive data to/from the end-user devices. The air interfaces are error prone. The specific applied parameters and algorithms applied to the air interfaces are listed in Table 28.

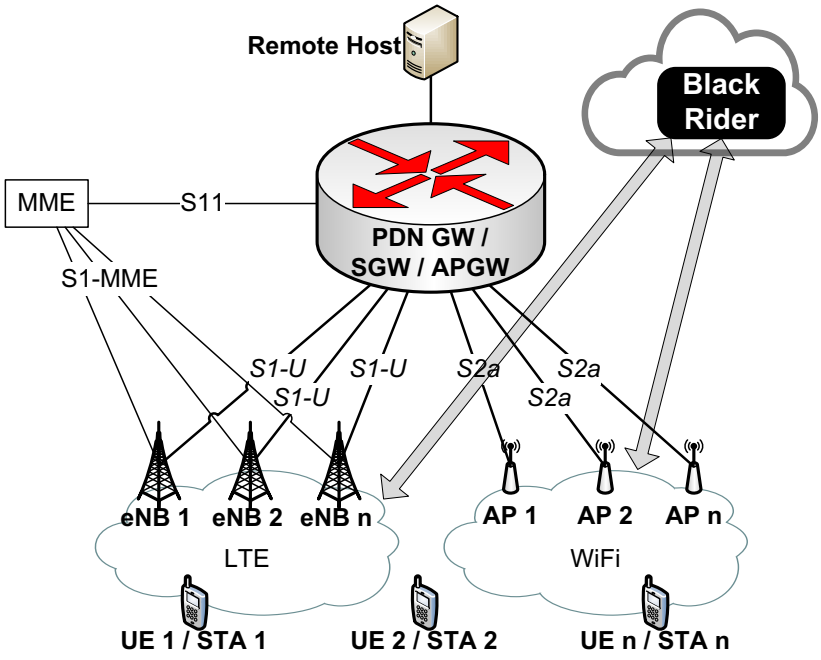


Figure 59: Simulation network architecture

8.2.1 Traffic Model

The end-user devices have each 3 dedicated bearers where UDP traffic is received and/or sent from/to the remote host. The first bearer consists of conversational video/voice traffic. The common transport protocol is UDP for video conferences. The second bearer consists of a You Tube video stream, which actually uses TCP as the transport protocol, but Google is developing a new UDP-based protocol called Quick UDP Internet Connections (QUIC) and therefore the YouTube traffic in the simulation is also based on UDP. The third bearer consists of an audio stream which uses the

UDP as the transport protocol as well. The first bearer is modelled as a bidirectional bearer, while bearer 2 and 3 are unidirectional bearers, since the data request sent in the uplink direction is very small in contrast to the data amount in the downlink direction and therefore the data in the uplink direction is neglected. In reality the packet size varies for all three bearers. In the simulation average packet size values have been identified and used as fixed packet sizes. For the bearer 1 it is 256 Bytes per packet according to (Szigeti and Hattingh, 2004). The bearer 2 uses a fixed packet size of twice the packet size of bearer 1 resulting in 512 Bytes. Bearer 3 has a packet size of 418 Bytes. This is calculated with equation (18) where the *SamplesPerFrame* is 1152, the *BitRate* is 128000 bps and the *SampleRate* is 44100Hz according to the standards (ISO/IEC 11172-3, 1993) (ISO/IEC 13818-3, 1998).

$$PacketSize = \frac{SamplesPerFrame}{8} * \frac{BitRate}{SampleRate} \quad (18)$$

It is assumed, that each frame is sent in a single packet. The traffic model is summarised in Table 27.

Table 27: Traffic model for the dedicated bearers

Bearer Number	Type of traffic	Bit Rate	UDP Payload Packet Size	Direction
1	Conversational video / voice	64 kbps (audio) + 320 kbps (video) = 384 kbps Audio codec e.g. G.711 (ITU-T, 1988), video codec e.g. H.264 (ITU-T, 2014)	256 Bytes	Uplink and Downlink
2	You Tube Video	770 kbps according to (Zink et al., 2008)	512 Bytes	Downlink
3	Audio Stream	128 kbps MP3 (ISO/IEC 11172-3, 1993) (ISO/IEC 13818-3, 1998)	418 Bytes	Downlink

The bit rates are constant for each bearer. Continuous streams are sufficient to analyse the effect of the BR on the offload use case.

8.2.2 Simulation Parameters

Table 28 shows the most important simulation parameters that are used within the simulation runs. The simulation parameters are grouped into general simulation parameters, LTE simulation parameters and Wi-Fi simulation parameters.

Table 28: Simulation parameters

General simulation parameters		
Parameter name	**Value**	**Description**
simTime	1000	Duration of the simulation in seconds [s]
useUdp	1	UDP is used as the transport protocol
endUserDeviceSpeed	3 Km/h 30 Km/h	Fixed speed of the end-user device in [Km/h].
mobilityModel	Random Walk	2D random walk mobility model. Each instance moves with a direction chosen at random and with fixed speed of 3 Km/h and 30 Km/h. If a boundary is hit, it is rebounded on the boundary with a reflexive angle.
LTE simulation parameters		
Parameter name	**Value**	**Description**
nMacroEnbSites	7	Number of macro cell sites (each site has 3 cells)
nMacroEnbSitesX	2	Minimum number of macro cell sites on the X-axis. They will be positioned in a 2-3-2 formation.
interSiteDistance	500	Distance between adjacent macro cell sites in metres [m]
macroEnbTxPowerDbm	46	Tx power for each macro cell in [dBm]
ueTxPowerDbm	10	UE Tx power in [dBm]
pathLossModel	Friis	Friis free space propagation model
physicalErrorModel	True	The simulator includes an error model according to the standard link-to-system mapping (LSM) techniques (Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), 2014).
scheduler	Proportional Fair	Proportional Fair Scheduler schedules a user when its instantaneous channel quality is high relative to its own average channel condition over time. (Sesia, Toufik and Baker, 2011).
macroUeDensity	0.00002	Number of UEs per square meter. This simulation setting results in 41 UEs.
macroEnbBandwidth	25	Number of resource blocks for each, uplink and downlink. 25 resource blocks lead to a 5 MHz DL and UL bandwidth.
handoverAlgorithm	A3Rsrp	A3Rsrp Handover Algorithm. The algorithm utilises Event A3 from the 3GPP standard (3GPP TS 36.331, 2014) section 5.5.4.4 with UE measurements and the Reference Signal Reference Power (RSRP). The event A3 is defined as the event when the UE perceives that a neighbour cell's RSRP is better than the serving cell's RSRP. The algorithm is verified and validated in (Budiarto et al., 2013) and (Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), 2014)
Wi-Fi simulation parameters		
Parameter name	**Value**	**Description**
nWi-FiSites	13	Numbers of Wi-Fi AP sites
nWi-FiSitesX	4	Minimum number of Wi-Fi AP sites on the X-axis of the Hotspot area.
interWi-FiSiteDistance	50	Distance between two nearby Wi-Fi AP cell sites in metres [m]
nWi-FiHotspots	12	A hotspot contains a group of APs. Number of hotspots distributed over the

		simulation area.
Wi-FiStandard	802.11g	The Wi-Fi IEEE standard that is used. The IEEE 802.11g standard is selected, because in the ns-3 the 802.11n standard is not completely implemented. There are no handover mechanisms implemented that improve the handover performance, such as the 802.11r standard (IEEE 802.11r, 2008).
pathLossModel	LogDistance	The Log-Distance Propagation Loss Model is appropriate to model the propagation loss inside a building or in densely populated area.
errorModel	Nist Error Model	An OFDM error rate for different modulations. The OFDM model description and validation can be found in (Pei and Henderson, 2010).
RTS/CTS	None	

Figure 61 shows the simulation area with the Signal to Interference plus Noise Ratio (SINR) values of LTE and the locations of the end-user devices (depicted as white dots with numbers) and APs forming the hotspots (depicted as black dots). The simulation area is given with 1500m on the X axis and 1300m on the Y axis. The simulation area represents a city area. The LTE simulation parameters from Table 28 lead to 7 sites, each of them includes a three sectorised eNB which realise together 21 macro cells. It is possible for a UE to move around the whole simulation area without losing the session (IP connection). A STA does not lose the IP session (IP connection) when moving around a hotspot are. A hotspot area consists of 13 APs and seamless session continuity can be guaranteed within the hotspot area and 25 m beyond the edge APs as depicted in Figure 60. The area in Figure 60 shows the area where session continuity can be guaranteed. It does not depict any signal quality. This area, where session continuity can be guaranteed, was identified through simulations with multiple STAs moving around with 3 Km/h and 30 Km/h applying the random walk mobility model. The simulation parameters given in Table 28 have been applied. It is possible to have session continuity beyond this area shown in Figure 60 due to the applied path loss model, but session continuity cannot be guaranteed outside this area.

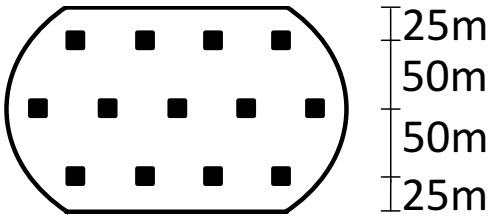


Figure 60: Guaranteed session continuity in the hotspot area

The UEs are randomly distributed within the simulation area, using a random uniform distribution along X and Y axis. The mode of LTE attachment is the idle mode cell selection (Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), 2014) which automatically connects the UE to an eNB. With the Wi-Fi simulation parameters from Table 28, there are 13 APs per hotspot and the 12 hotspots are randomly distributed using a random uniform distribution along X and Y axis over the simulation area. The numbers of LTE macro cells and Wi-Fi hotspots form a city area where the overall network coverage is provided by LTE macro cells and in specific areas the Wi-Fi antennas are grouped to Wi-Fi hotspot in order to offer additional coverage and capacity to the end-user devices. The macroUeDensity parameter from Table 28 with a value of 0.00002 results in 41 end-user devices in this simulation area.

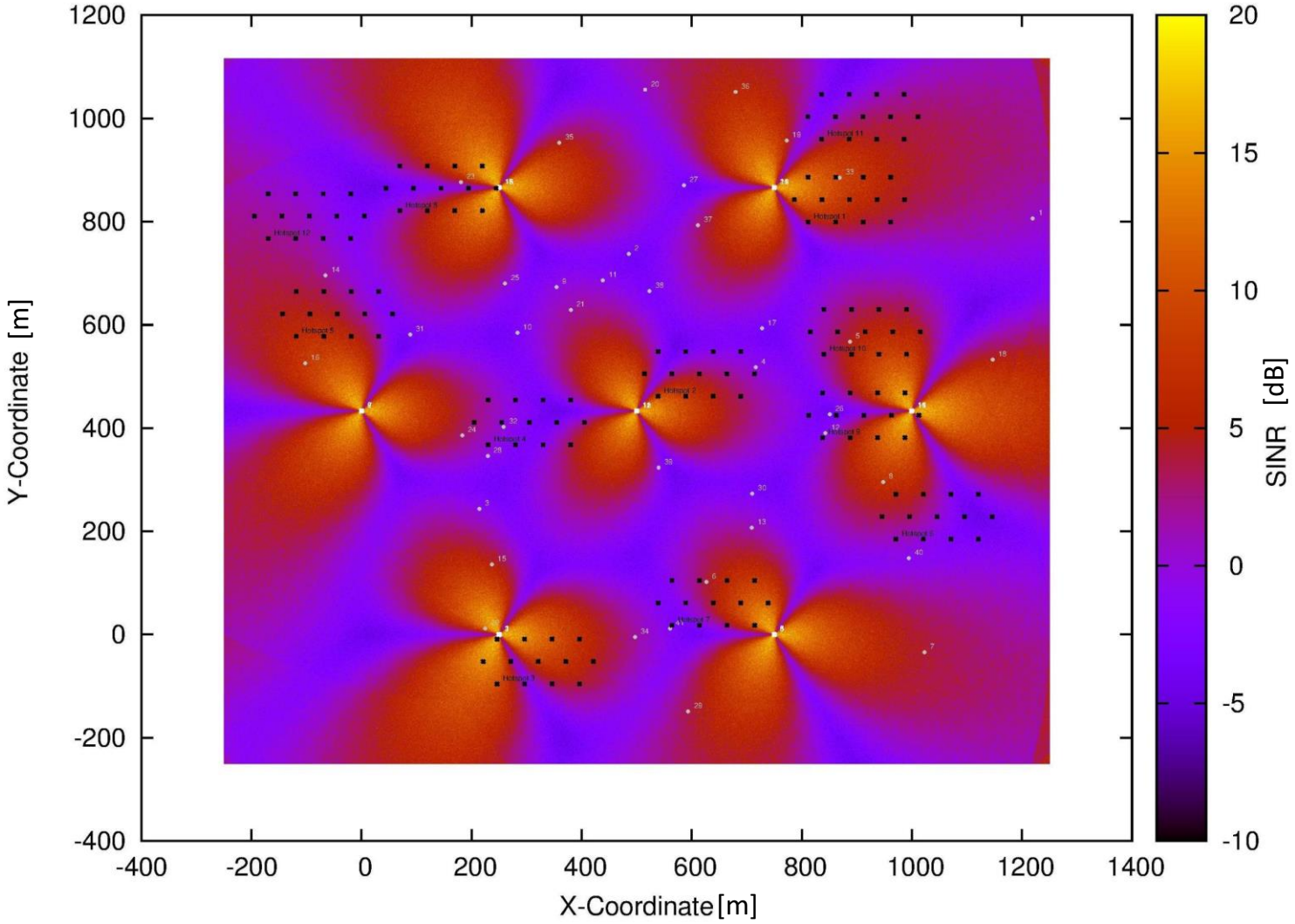


Figure 61: Simulation area with LTE SINR values and locations of hotspots, eNBs and end-user devices

8.2.3 Zero Simulation Scenario

To enable the comparison of the simulation scenarios with and without the use of the BR, one zero simulation variant is introduced. The zero simulation variant is based on the variant 0 of section 8.1.1 of the traffic steering scenarios described in section 8.1. It represents the simulation scenario considering how offloading and inter-system handover work nowadays. In the following, the zero simulation scenario is defined:

- Zero simulation scenario: 100% of the users do have a CM installed, which offloads bearer 2 and 3 whenever the end-user device is associated with an AP. The abbreviation of this scenario is **zero_100_Xkmph**. The X is substituted with either 3 or 30 which indicates the speed of the end-user device.

For every simulation scenario with the BR this zero simulation scenario is used to compare the BR simulation scenario to it. It is used as reference simulation scenario.

8.2.4 Black Rider Simulation Scenarios

The simulation scenarios using a BR are based on the variant 2 of section 8.1.3 which results in a network-controlled, terminal assisted mode where the PoD is located at the network. As a result of the selected variant 2, the BR sends strict commands to the end-user devices which are executed promptly at the end-user devices.

There are two BR simulation scenarios defined which are introduced in the following:

1. BR simulation scenario 1: All end-user devices apply the BR. The BR enforces policies by considering the location of the end-user device. The end-user devices are only allowed to associate with a Wi-Fi AP, if they are in the location area of a hotspot where session continuity can be guaranteed, as shown in Figure 60. The threshold value is given in the simulation with 25m, which leads only to an association with an AP if an end-user device is less or equal 25m away from an AP at the edge of a Wi-Fi hotspot. Figure 60 shows the guaranteed session continuity area within a hotspot. As soon as an end-user device is associated with an AP, bearer 2 and bearer 3 are offloaded to Wi-Fi. The abbreviation for this simulation scenario is **BR_Location_Xkmph**.
2. BR simulation scenario 2: All end-user devices apply the BR. The BR analyses the history of the end-user device. This is done by analysing the end-user

device movement, as well as the received packets related to the end-user device location of the zero_100 simulation scenario runs. As a result, the predictions of the end-user device movements are precise, because the usage of the same seed lead to the same movements of the end-user devices. As soon as an end-user device is associated with an AP, the bearer 2 and bearer 3 are offloaded to Wi-Fi. Furthermore, the reception of packets at the end-user device is logged. Based on this data, the throughput is analysed and offloading to Wi-Fi is only allowed if an average throughput per offload of more than 300 kbps for bearer 2 and more than 49 kbps for bearer 3 is achieved. This simulation scenario concatenates two external modules, the mobility module, analysing the received packets and location and the throughput module, analysing the throughput. The abbreviation for this simulation scenario is **BR_History_Xkmph**.

To achieve the functionality of the BR_History simulation scenario in reality, this scenario would contain the application of a mobility analyser module, which identifies stable places where the end-user device moves only little, such as at home, at the working place, at the university etc. Based on this information the mobility analyser is able to make predictions of the end-user device movement. The other external module would be a throughput analyser or a packet receiving analyser. The throughput analyser logs all the throughputs and the corresponding locations and the associated WLAN (SSID). With this information it is possible to decide if an offload is suitable or not, regarding the throughput values. The packet receiving analyser could collect the end-user device movements and the received packets per location over a long time. With this combination of information it is possible to create a map such as depicted in Figure 62. The area could be divided into small subareas and when a certain density of received packets at the end-user device is reached, traffic offload to Wi-Fi is allowed in this small area. These maps may be also associated with the end-user device speed where offloading is permitted if the end-user device speed is higher than 30 Km/h, because Wi-Fi is not suitable for high speeds.

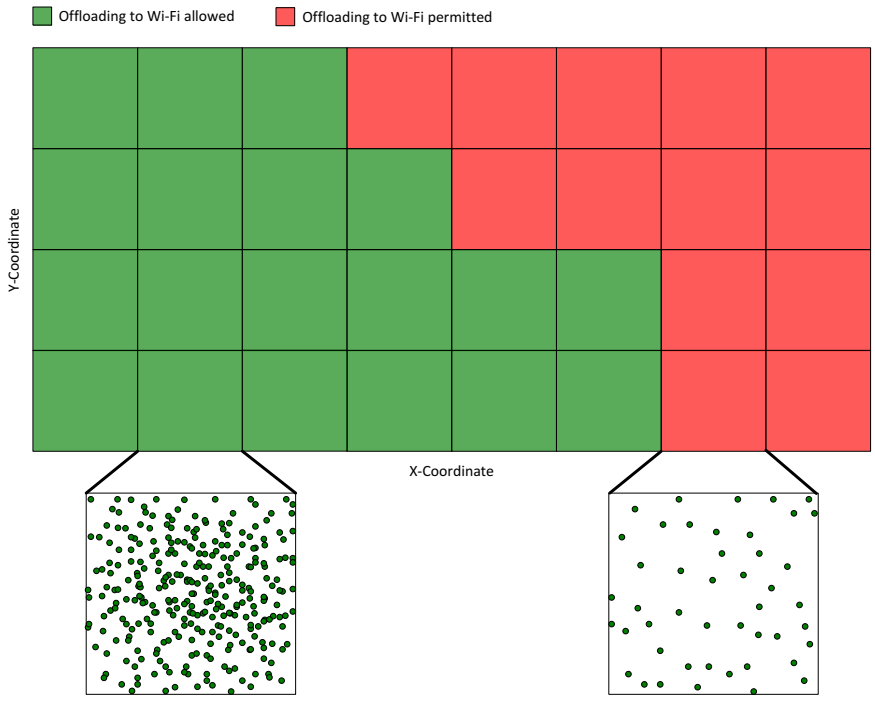


Figure 62: BR history-based offload permission map

This BR_History scenario could be applied with and without a concatenation of a mobility analyser module. Without a mobility analyser, an additional policy could be applied so that the small area has to have a certain number of small neighbour areas that also allow offloading to Wi-Fi. This would reduce the risk of very short offloading times to Wi-Fi.

8.2.5 Types of Simulation Results

The types of simulation results are the same for every simulation scenario and are listed in the following:

- Histogram of number of offloads. The number of offloads is counted the following way: Whenever an end-user device was associated with LTE and then offloads bearer 2 and bearer 3 traffic to Wi-Fi, this is counted as one offload. There are three kinds of offload numbers that are counted or calculated in the simulation. The total numbers of offloads, the number of offloads where data traffic has been received by the end-user device, and the number of offloads where no traffic has been received. The latter represents the worst offload case, because an offload is performed, but this was unnecessary, due to no receiving of data was possible. The cause for an offload, where no data have been received by the end-user device, is a poor signal

quality due to movement of the end-user device and path loss applied in the simulation.

- Total number of received Kilo Bytes (KB) for the whole simulation time on the downlink separated by the 3 bearers and by the access technologies. The received bytes on the downlink consist of data traffic only. The signalling traffic is not considered.
- Average throughput in kbps on the downlink separated by the 3 bearers and by the access technologies. The average throughputs are calculated from the point in time when the end-user device is associated with a RAT to the point in time when the RAT is changed. This time period, where the throughput is calculated, is shown in Figure 63 with green lines. The average throughputs are calculated for each RAT and each bearer per association to the specific RAT. The average throughputs are then summarised over each association to the specific RAT for each bearer and the overall average is calculated. Figure 63 shows an example of the bearer 2 which is offloaded to Wi-Fi.

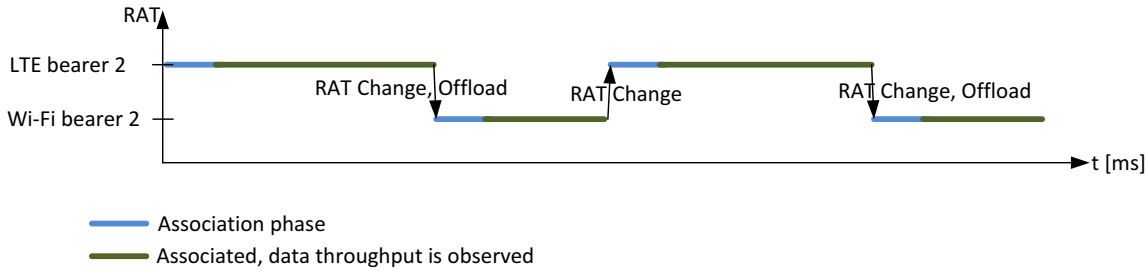


Figure 63: Throughput calculations

Equation (19) is used to calculate the average throughput per bearer and per RAT. The variable x represents the bearer number, which is either 1, 2, or 3. The RAT is either LTE or Wi-Fi. The variable n represents the number of associations per RAT and per bearer.

$$AverageThroughputBearer_xRAT = \frac{1}{n}\sum_{i=1}^{n} Bearer_xRATthroughput_i \qquad (19)$$

These types of results enable conclusions to be drawn if the BR can improve these values or not with the offload use case. The simulation results of the simulation scenarios are presented and discussed in section 8.3.

8.2.6 Steady State

The steady state of the simulation is reached when the network architecture is set up and the applications, which generate the traffic, are running. This period of time is also called the warming-up period. In the simulation, the steady state is reached 0.013 seconds after starting the simulation. The simulation data is collected after the 0.013 seconds. Simulation data before 0.013 seconds are not considered and therefore discarded.

8.2.7 Simulation Plan

Every simulation scenario is executed with 2 different speeds: 3 and 30 Km/h. This results in the simulation plan in Table 29. The simulation contains 1 simulation run per simulation scenario. Every end-user device has its own seed for the mobility algorithm.

Table 29: Simulation plan

Simulation Scenario	Seed	Simulation Duration	Number of Simulation Runs
Zero_100_3kmph	1	1000 Seconds	1
BR_Location_3kmph	1	1000 Seconds	1
BR_History_3kmph	1	1000 Seconds	1
Zero_100_30kmph	1	1000 Seconds	1
BR_Location_30kmph	1	1000 Seconds	1
BR_History_30kmph	1	1000 Seconds	1
Total Number of Simulation Runs			**6**

8.2.8 Confidence Interval

To calculate the confidence interval for the observed data, the sample mean is calculated using equation (19), where N is the numbers of values collected.

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (20)$$

The sample variance is an unbiased estimate of the population variance and is calculated with equation (21).

$$S^2{}_{N-1} = \frac{1}{N-1}\sum_{i=1}^{N} (X_i - \overline{X})^2 \qquad (21)$$

Since the sample mean and the sample variance are unknown the student-t derivation is applied. Therefore, the confidence interval is calculated with equation (22) where the confidence level is 95%.

$$\overline{X} \pm t_{\left(\frac{1-\alpha}{2},N-1\right)} \frac{\sqrt{S^2}}{\sqrt{N}} \qquad (22)$$

The confidence intervals per end-user device speed, per result type, and per simulation scenario are shown in Table 30 and Table 31. All the results are average values over all the end-user devices of the simulation. The throughput is given in kbps whereas the received amount in the downlink is given in KB. The confidence intervals of BR _Location and BR_History are for most of the values smaller than the confidence intervals of zero_100 simulation scenario. This leads to the conclusion that the behaviour of the different end-user devices is more regular and balanced with the offload use case with the use of the BR than without.

Table 30: Confidence interval for all simulation scenarios for 3 Km/h

	Zero_100	BR_Location	BR_History
ReceivedKBLteBearer1	44375 ± 2396	43875 ± 2426	44500 ± 2391
ReceivedKBLteBearer2	49254 ± 5171	57858 ± 2956	56368 ± 2818
ReceivedKBLteBearer3	7698 ± 846	9308 ± 484	9252 ± 471
ReceivedKBWi-FiBearer2	32180 ± 3539	30431 ± 1578	32130 ± 1551
ReceivedKBWi-FiBearer3	5159 ± 557	4878 ± 253	5150 ± 248
AverageThroughputkbpsLteBearer1	355 ± 18.8	351 ± 18.9	356 ± 18.5
AverageThroughputkbpsLteBearer2	723 ± 38.3	721 ± 38.9	725 ± 38.5
AverageThroughputkbpsLteBearer3	113 ± 5.8	116 ± 6	119 ± 5.7
AverageThroughputkbpsWi-FiBearer2	565 ± 57	680 ± 41	698 ± 34
AverageThroughputkbpsWi-FiBearer3	90 ± 9.3	109 ± 5.7	114 ± 5.3
NoOfOffloads	2.9 0.3	2 ±0.09	2.1 ± 0.08
NoOfOffloadsWithoutRx	0.3 ± 0.04	0 ± 0	0 ± 0

Table 31: Confidence interval for all simulation scenarios for 30 Km/h

	Zero_100	BR_Location	BR_History
ReceivedKBLteBearer1	43625 ± 2402	43375 ± 2446	44000 ± 2411
ReceivedKBLteBearer2	51444 ± 5509	61009 ± 3180	59874 ± 3053
ReceivedKBLteBearer3	8442 ± 947	9717 ± 515	9675 ± 503
ReceivedKBWi-FiBearer2	28854 ± 3237	27405 ± 1450	28764 ± 1417
ReceivedKBWi-FiBearer3	4626 ±509	4392 ± 232	4610 ± 227
AverageThroughputkbpsLteBearer1	349 ± 18.8	347 ± 19.1	352 ± 18.6
AverageThroughputkbpsLteBearer2	719 ± 38.8	722 ± 39.7	724 ± 39
AverageThroughputkbpsLteBearer3	118 ± 6.3	115 ± 6	117 ± 5.7
AverageThroughputkbpsWi-FiBearer2	539 ± 55	676 ± 41	692 ± 35
AverageThroughputkbpsWi-FiBearer3	86 ± 9	108 ± 5.8	112 ±5.6
NoOfOffloads	29.3 ± 3.8	18 ± 0.8	18.8 ± 0.7
NoOfOffloadsWithoutRx	3.8 ± 0.48	0 ± 0	0 ± 0

8.3 Simulation Results and Evaluation

The ability of implementing the BR architecture has shown that the BR architecture is coherently executable and implementable. In the following, the gained results from the defined simulation plan in 8.2.7 are presented and discussed.

8.3.1 Offload Numbers

In the following the results of the offload numbers are analysed for the two different speeds, 3 Km/h and 30 Km/h. The numbers of offloads without received data are wasted offloads, because an offload is performed, but no data could have been received by the end-user device. The time duration when an end-user device is associated with an AP, but is unable to receive any data over this connection is called idle association time in this thesis. It is important to reduce this idle association time as much as possible to enable an efficient use of the Wi-Fi RAT.

Figure 64 a) and Figure 64 b) show the simulation results of the offload numbers for 3 Km/h and 30 Km/h end-user device speed. The BR_Location and the BR_History simulation scenarios both have no unsuccessful offloads with both speeds. Whenever an offload is performed with the BR simulation scenarios it is possible for the end-user device to receive data. Therefore, the idle association time is 0. The reduction of the idle association time to 0 is possible because with the BR_Location simulation scenario offloading is only allowed when the end-user device is located in the guaranteed session continuity area within a hotspot, see Figure 60. With the BR_History simulation scenario it is possible to reduce the idle association scenario with the applied external mobility analyser and the throughput modules. These two external modules only allow offloads when packets could have been received with the zero_100 simulation scenario and the average throughput is greater than 300 kbps. The zero_100 simulation scenario shows a number of unsuccessful offloads where no data reception in the downlink is possible. The unsuccessful offloads are caused due to associations with edge APs where the signal strength is initially sufficient though low. The association to the AP takes place but then the signal strength is that low that it is impossible to receive any data at the end-user device. Such a decrease in signal strength can occur because of the path loss model applied for the Wi-Fi signal propagation. This does not cause a continuous omni-directional way of signal

degradation, instead the signal strength gets more irregular, the further away from the antenna the signal is received.

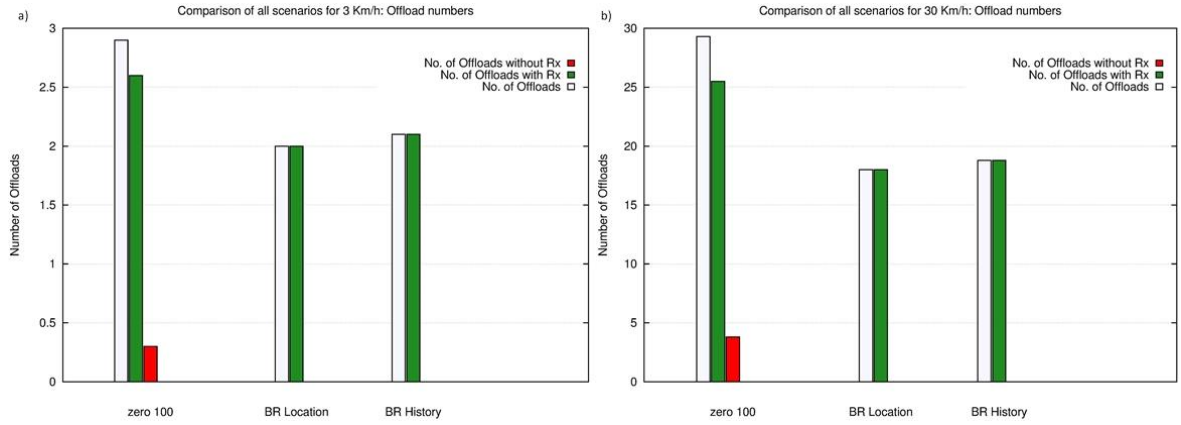


Figure 64: Comparison of offload numbers for all scenarios for 3 and 30 Km/h

An unnecessary offload, where an offload to Wi-Fi is performed, but the end-user device is not able to receive any data leads also to a waste of energy, because during this idle association time the Wi-Fi interface at the end-user device is fully up and running without being able to receive any data.

Figure 65 shows the percentage of successful offloads to Wi-Fi. The percentage of successful offloads is higher with the zero_100 simulation scenario with 3 Km/h than with 30 Km/h. The possibility that the end-user device is unable to receive any data due to low signal is higher with an end-user device speed of 30 Km/h than with 3 Km/h. When the BR is applied, the percentage of successful offloads is 100%. The reason for this 100% offloading success rate with the BR is because it is only allowed to associate with an AP if the end-user device is located within the guaranteed session continuity area within a hotspot in case of the BR_Location simulation scenario. Within the guaranteed session continuity area within a hotspot it is always possible to receive data because the signal quality is sufficient. With the BR_History simulation scenario the zero_100 simulation scenario is analysed regarding the end-user device movement, the received packets related to the end-user device location, and the achieved average throughput. This enables the BR to command the end-user devices in a way that no unsuccessful offloads occur. In reality, such a precise prediction is not possible. Therefore, these results have to be treated with care due to the exact manner.

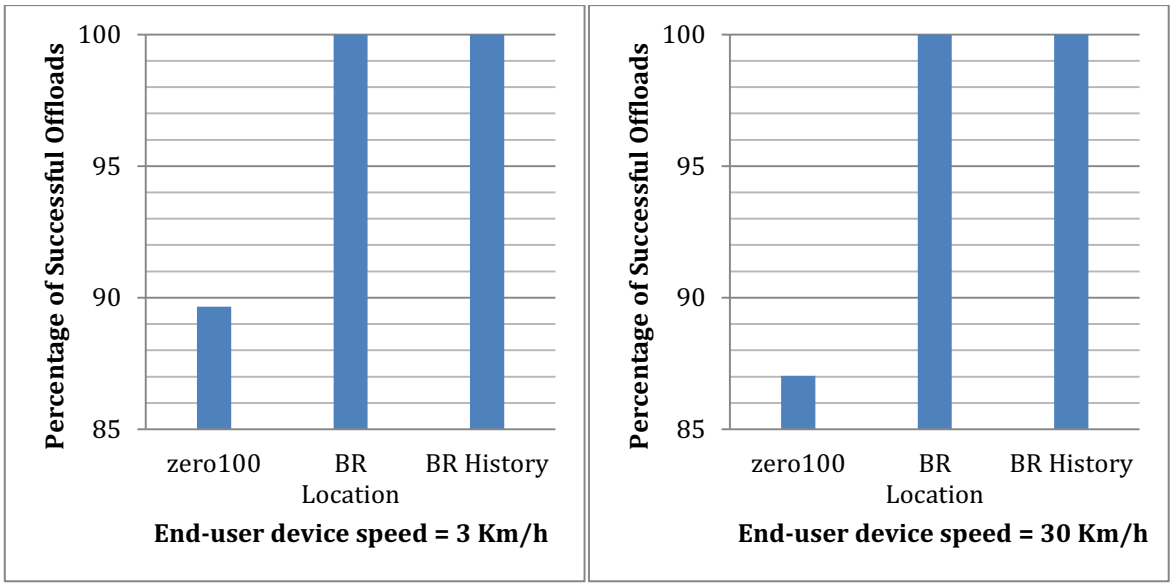


Figure 65: Percentage of successful offloads per simulation scenario and speed

The simulation results showed that the application of the BR can eliminate unnecessary offloads and therefore eliminate the idle association time for both end-user device speeds.

8.3.2 Received KB in Downlink Direction of LTE and Wi-Fi Bearers

In the following the simulation results of the received Kilo Bytes (KB) in the downlink direction (received by the end-user devices) for the whole simulation time per bearer and per RAT for the two different speeds, 3 Km/h and 30 Km/h are analysed.

An interesting value is the idle association time. The lower the idle association time is, the higher the value of the received KB on LTE gets. During an idle association time the end-user device is connected over Wi-Fi and is not able to receive any data. If the idle association time is eliminated, the end-user device is connected during this time over LTE and is able to receive data. Therefore, the KB value on LTE is increased in this case. Since there is no offloading on bearer 1, the received KB values are very similar for the different simulation scenarios and therefore bearer 1 is not considered in this analysis. As mentioned before, the BR_Location and the BR_History are idealised simulation scenarios and the values have to be treated with care.

Figure 66 a) and Figure 66 b) show the simulation results of the received KB values for the whole simulation time per simulation scenario, per RAT, and per bearer for 3 Km/h and 30 Km/h end-user device speed. The received KB values on LTE of the

BR_Location simulation scenario are the highest of all the simulation scenarios for both end-user device speeds, because it has the lowest number of offloads and therefore most of the time the bearers are connected through LTE. The second highest value of received KB values on LTE for both end-user device speeds is achieved with the BR_History, which is only slightly lower than the values from the BR_Location simulation scenario. With a notable lower value of received KB on LTE for both speeds follows the zero_100 simulation scenario. These lower values of received KB are caused through the idle association time, where no data traffic could be received during offloading. Since the BR simulation scenarios both eliminate the idle association time, it is possible for the BR simulation scenarios to send over LTE during this time while no data is received with the zero_100 simulation scenario.

The values of received KB for Wi-Fi show the highest values with the zero_100 simulation scenario followed by the BR_History and the BR_Location simulation scenario for both end-user device speeds. The highest value of received KB is achieved with the zero_100 simulation because whenever a Wi-Fi signal is received, an association with the AP is performed. This results in the highest values of received KB. The values of received KB with the BR_Location for Wi-Fi are the lowest because an association with an AP is only allowed if the end-user device is located in the guaranteed session continuity area within a hotspot. Therefore, the number of performed offloads is lower than with the zero_100 and the BR_History simulation scenario. Since the BR_History simulation scenario is based on the analysis of the zero_100 simulation scenario, these values are very similar to each other, but still the BR_History values of the received KB are slightly lower than the values from the zero_100 simulation scenario. The reason for this is that with the BR_History simulation scenario offloading is only allowed if the average throughput is greater than 300 kbps and therefore less offloads are performed. As a result, less KB is received over Wi-Fi at the end-user devices. The values of the received KB from the BR_History simulation scenario are slightly higher than the values from the BR_Location because the offload decision is not only based on the location. If data could have been received with the zero_100 simulation scenario, and the average throughput was more than 300 kbps, the BR allows offloading traffic regardless of the distance between the end-user device and the edge AP.

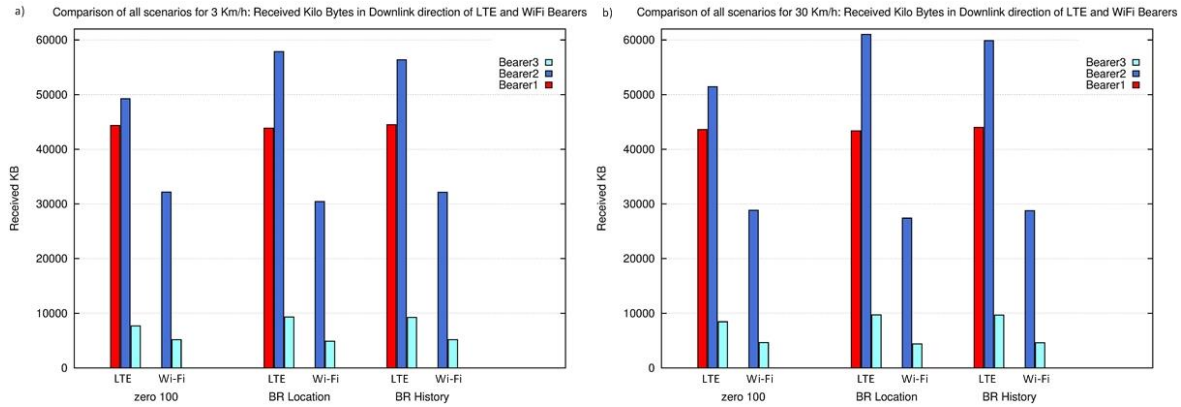


Figure 66: Comparison of Received KB in downlink direction of LTE and Wi-Fi bearers for all scenarios

The summarised values of the received KB per bearer for LTE and Wi-Fi for the whole simulation time from the zero_100, the BR_Location, and the BR_History simulation scenarios are used for comparison. The received KB average values are calculated for the bearer 2 and 3. The results are shown in Figure 67. It shows clearly that the summarized received KB values for bearer 2 and bearer 3 are higher if the BR is applied.

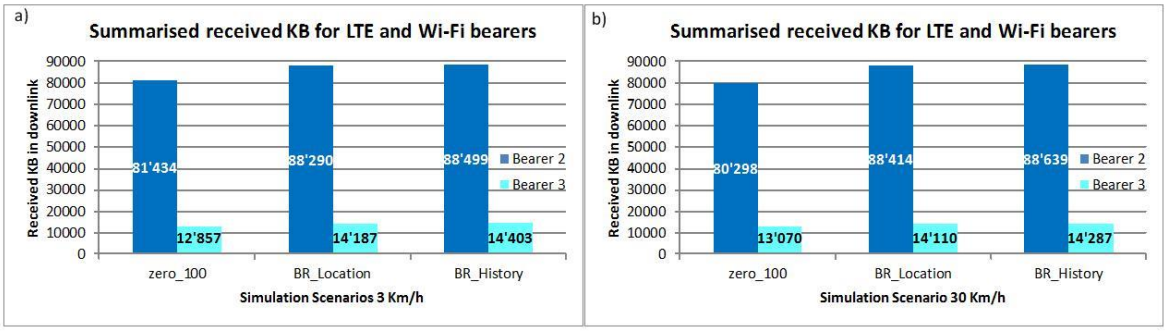


Figure 67: Summarized received KB for LTE and Wi-Fi bearers for 3 and 30 Km/h

Table 32 shows the percentage improvement of the overall received KB values throughout the whole simulation time per bearer and per end-user device speeds against the zero_100 simulation scenario. When the BR is applied 7.9 to 12 percent more KB could be received in the downlink through the whole simulation time per bearer. The variation of the percentage values are caused due to the dispersion of the simulation results.

Table 32: Percentage improvement of the overall received KB values against the zero_100 simulation scenario

3 Km/h	Bearer	BR_Location	BR_History
% improvement of the overall received KB value against the zero_100 simulation scenario	2	8.4	8.6
	3	10.3	12
30 Km/h	**Bearer**	**BR_Location**	**BR_History**
% improvement of the overall received KB value against the zero_100 simulation scenario	2	10.1	10.3
	3	7.9	9.3

The simulation showed that the end-user devices served by a BR can receive more KB in summary for the whole simulation time as without the application of the BR, because the idle association time is reduced. This is either achieved with the decision not to offload data traffic to Wi-Fi or to return earlier to LTE than in the case where the BR is not applied.

8.3.3 Throughput in Downlink Direction of LTE and Wi-Fi Bearers

In the following the simulation results of the throughput values in downlink direction per bearer and per RAT for the two different speeds, 3 Km/h and 30 Km/h are analysed. Data from bearer 1 is not offloaded to Wi-Fi and therefore the throughput values are similar for all the simulation scenarios. The throughput values for bearer 2 and 3 on LTE are similar, because on LTE there is no idle association time and therefore only slight differences occur on the throughput. As a result, only the Wi-Fi throughput values are analysed in the following.

Figure 68 a) and Figure 68 b) show the simulation results of the throughput values per simulation scenario, per RAT, and per bearer for 3 Km/h and 30 Km/h end-user device speed. For both end-user device speeds the highest Wi-Fi throughputs are achieved with the BR_History simulation scenario followed by the BR_Location and the zero_100 simulation scenarios. Compared to the end-user device speed of 3 Km/h the throughput of all Wi-Fi bearers are lower with 30 Km/h. The biggest difference between the Wi-Fi throughput between 3 Km/h and 30 Km/h occurs with the zero_100 simulation scenario. With the application of the BR simulation scenarios it is possible to get better throughput results even with the higher end-user device speed than with the zero_100 simulation scenario. The reason for these results is the elimination of the idle association time and the slightly better throughput results of the BR_History compared to the BR_Location is caused through the analysis of the received packets and the corresponding end-user device location and the fact that the

throughput is analysed and offloading is only commanded by the BR if the average throughput is greater than 300 kbps.

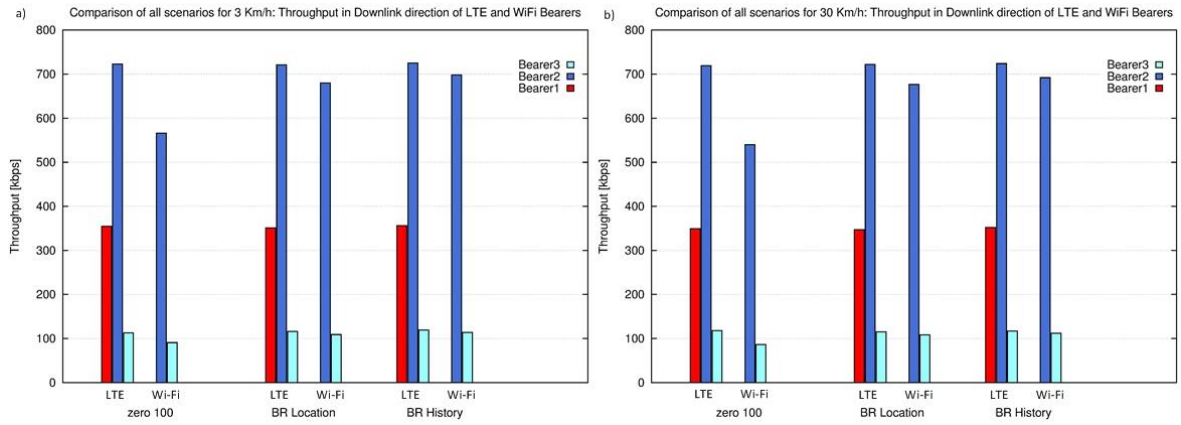


Figure 68: Comparison of throughput in downlink direction of LTE and Wi-Fi bearers of all scenarios

Table 33 shows the percentage improvement of the average throughput Wi-Fi values per bearer and per end-user device speeds against the zero_100 simulation scenario. Only the throughput values on Wi-Fi are considered, because the LTE throughput values are very similar with each simulation scenario. The average throughput on Wi-Fi can be increased between 20 to 29 percent, which is a massive gain compared with the average Wi-Fi throughput without the BR. The reason for this gain is the elimination of the idle association time, that offloading is only allowed when in the area where session continuity is guaranteed (Figure 60), and that the BR_History only allows offloading if packets have been received by the end-user device in the zero_100 simulation scenario and the average throughput is greater than 300 kbps.

Table 33: Percentage improvement of the average Wi-Fi throughput values against the zero_100 simulation scenario

3 Km/h	**Bearer**	**BR_Location**	**BR_History**
% improvement of the average Wi-Fi throughput value against the zero_100 simulation scenario	2	20	23
	3	20	25
30 Km/h	**Bearer**	**BR_Location**	**BR_History**
% improvement of the average Wi-Fi throughput value against the zero_100 simulation scenario	2	25	28
	3	25	29

The number of offloads are analysed regarding the average throughput of the bearer 2 for the end-user device speed 3 Km/h and 30 Km/h. Therefore, 5 classes of average throughputs are defined, one average throughput value consists of 0 kbps and 4 average throughput ranges consisting of the following throughputs: 0 kbps < w ≤ 100 kbps, 100 kbps < x ≤ 200 kbps, 200 kbps < y ≤ 300 kbps, and 300 kbps < z. The first 4 throughput classes contain all together the range from 0-300 kbps because this

average throughput range is considered as low for the bearer 2 which streams a you tube video with 770 kbps. This range is shown with red shades in Figure 69 and Figure 70. The 5th throughput class, containing average throughputs greater than 300 kbps are considered as acceptable and therefore grouped together. This range is shown in green in Figure 69 and Figure 70. Figure 69 and Figure 70, containing the results of 3 Km/h and 30 Km/h end-user device speeds, are compared with each other and discussed in the following. The zero_100 simulation scenario has a red shaded (0-300 kbps) range of 27.5 % with 3 Km/h and 35.8 % with 30 Km/h. The percentage of the red shaded range with 30 Km/h is higher than with 3 Km/h because of the higher end-user device speed. If the end-user device moves with a higher speed, the decrease of the signal quality takes place faster than with a lower speed. Therefore, the throughput is decreasing more often with 30 Km/h than with 3 Km/h. The highest percentage value of the red shaded range occurs with 0 kbps, where no data could be received by the end-user device. With the BR_Location simulation scenario the first two average throughput classes can be eliminated, because the BR only commands an offload if the end-user device is located within the guaranteed session continuity area within a hotspot. The lowest average throughput within this area is between 100 and 200 kbps, but not below. The percentage of the two red shaded throughputs (100-300 kbps) is with 0.55 % for 3 Km/h and with 0.48 % for 30 Km/h very low. The fact, that the red shaded throughputs with 30 Km/h end-user device speed is lower than the one with 3 Km/h end-user device speed is explainable with the dispersion of the simulation results. Whenever an end-user device is entering a guaranteed session continuity area within a hotspot, the BR commands an offload and whenever an end-user device is leaving this area, the BR commands an inter-system handover to LTE. Therefore, the behaviour is similar for both end-user device speeds. The BR_History simulation scenario has a percentage of average throughputs over 300 kbps of 100 %. This result is achieved through the applied external throughput and mobility analyser modules with this simulation scenario, where the mobility and the received packets and the associated location from the zero_100 simulation scenario are analysed and offloads are only allowed if the average throughputs are greater than 300 kbps. The difference between the BR_Location and the BR_History simulation scenario is very small because the

average throughput in the guaranteed session continuity area within a hotspot is very good.

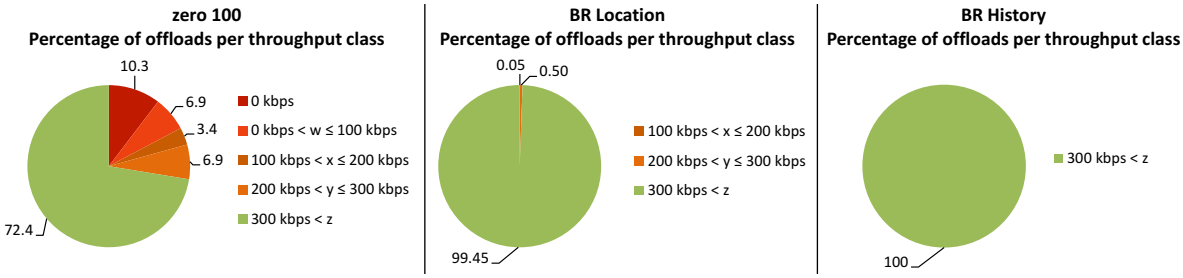


Figure 69: Percentage of offloads per throughput class per simulation scenario with 3 Km/h for bearer 2

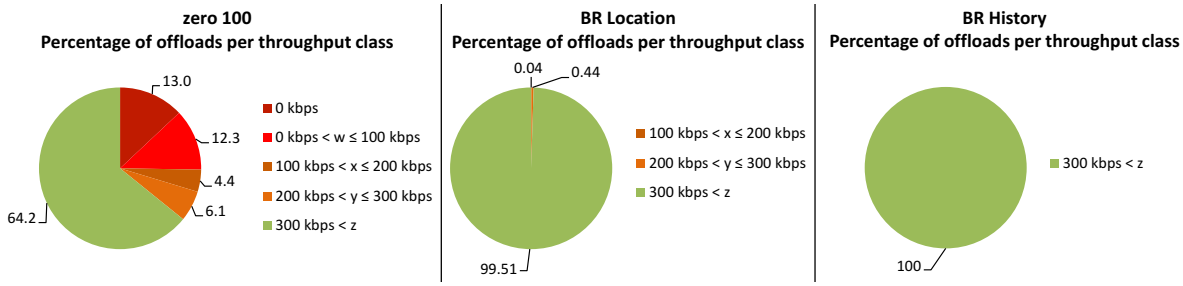


Figure 70: Percentage of offloads per throughput class per simulation scenario with 30 Km/h for bearer 2

The simulation showed that the throughput, when applying the BR, is clearly better on Wi-Fi than without the application of the BR. This improvement of the throughput with the BR_Location simulation scenario is realised through the elimination of the idle association time with the decision of not offloading data traffic to Wi-Fi when not located in the guaranteed session continuity area within a hotspot and returning earlier to LTE than in the case where the BR is not applied. An earlier return from Wi-Fi to LTE prevents the end-user device from getting a worse and worse signal quality which is causing a decrease of the throughput. Furthermore, with the BR_History the application of the external mobility analyser and throughput modules improve the throughput even more and eliminates the idle association time as well. With the application of the BR the low throughputs from 0-300 kbps can be nearly eliminated (BR_Location) or even completely eliminated (BR_History).

8.3.4 Discussion

The simulation results showed that the BR is able to significantly improve the traffic steering with intelligent offload commands that lead to an elimination of the idle association time and as a result of this, the throughput and the summarised received data for the whole simulation time over Wi-Fi and LTE are increased, and the

unnecessary energy consumption is reduced. The ability of the BR to concatenate and coordinate external modules is prooven with the BR_History simulation scenario by coordinate the mobility analyser and the throughput module. The afore mentioned BR use case, to eliminate the need of adding an offset to the SSDL to prevent an over-utilisation of macro cells described in section 7.4, is as well proved to be feasible through the BR_Location simulation scenario, which offloads traffic based on the location to small cells.

The simulations showed that the results of received KB for the whole simulation time, the throughput, and the offload numbers, where receiving of data is possible, are throughout better if the BR is applied. As a result of these improved values with the BR, another use case, where the BR could be applied and improve the performance, is the LTE/Wi-Fi link aggregation. The LTE/Wi-Fi link aggregation enables the combination of the LTE and Wi-Fi RATs for the same session. This means that both RATs could be used in parallel to download, for example, a video. A situation where Wi-Fi performs badly is if an association to an AP is possible, but due to the bad signal quality only few or no data can be sent and received over the Wi-Fi link. In such cases, where Wi-Fi does not bring many improvements to the end-user device, the BR is able to eliminate these cases, as could be seen with the analysis of the percentage of offloads per throughput class and per simulation scenario. As a result, the end-user device saves energy and the LTE/Wi-Fi link aggregation is only applied if it is reasonable. With the introduction of the aforementioned BR history-based offload permission map showed in Figure 62, the identification of appropriate and inappropriate locations for the usage of Wi-Fi is possible. As a result of the application of the BR history-based offload permission map, the LTE-Wi-Fi link aggregation use case can be improved as well.

Even if the simulation shows only a limited area of the possibilities with the BR due to the limited numbers of considered parameters used in the simulation, it is proved that the BR can provide benefits for several use cases. This gives an impression of how much more benefit the BR could bring, if more parameters would be considered to be included in the traffic steering process.

9 Conclusions and Directions of Further Research

To conclude this thesis, the achievements and limitations of the thesis are summarised in this chapter. The chapter proceeds with considerations about possible directions of further research based on the results of the presented research results.

9.1 Achievements of the Research

The massive increase of data traffic volume of the last years and the continuation of this development in future years bring up the need for a holistic approach to tackle the resulting and arising problems and demands of the heterogeneous mobile networks. The issues that arise are, for example, the improvements of QoS, the increasing end-user device mobility when being always on, and the user demand to be always best connected etc. The mobile network environment is a very complex one and the change to heterogeneous mobile networks makes it even more complex. This research aimed to define and evaluate an advanced network architecture that is able to provide a feasible and modular basis to provide novel traffic steering possibilities in heterogeneous mobile networks to tackle the problem of the enormous traffic volume while improving the QoS, support seamless mobility, and provide valuable services to the end-user.

The existing architecture has been analysed and based on this and further analysis on the architecture details have led to the basic design selections in chapter 4 for the proposed network architecture. The design selections have the aim of narrowing the highly complex environment of mobile networks, to focus on specific areas within the mobile network architecture, and to provide the basis of a feasible integration of the proposed architecture into the existing one.

The Black Rider (BR) has first been introduced on a high level in chapter 5 to provide afterwards a detailed definition of the BR architecture, its building blocks, and the functions and processes between them in chapter 6. The BR architecture has been defined on the basis of the design selection focusing on a feasible integration into the existing architecture by re-using existing protocols that can be applied to provide smart, intelligent and individual traffic steering. The data gathering is realised with

several different protocols to enable the data gathering on the LTE 3GPP access and on the Wi-Fi access to collect the context data of the end-user device as well as of the network state and utilisation. The feasibility of integrating the BR into existing mobile networks architectures is additionally supported by the external modules, which can be operator defined as well as 3rd party defined. The integration of new external modules is feasibly done through an API definition and the BR database (DB) access is provided through the nwI_BR DB interface. By placing the individual BRs at the cloud, a dynamic way of deployment is enabled and the delay of the traffic between the BR and the end-user device can be reduced due to a location change of the BR if the geographic distance to the end-user device gets too high and another datacentre is placed nearer to the end-user device. The BR is real-time capable of distributing the commands and information to the end-user device.

The BR can be exploited to gain benefits in multiple use cases which are described in chapter 7. With the use of the BR it is possible to provide individual control information over the BR individual control channel instead of sending common control information which is less exact to the end-user devices in case of handover and offload situations. The use of the BR and therefore a smart handover and offload behaviour can result in energy savings at the end-user device, as shown with the calculations of the energy efficient use case. The provision of much more exact geolocation information from the BR to 3rd parties generates a massive additional value and benefit to 3rd parties using this information to provide more exact services such as for CDNs, site-specific advertisements etc. The problem of under-utilisation of small cells and the over-utilisation of macro cells in heterogeneous mobile networks is solved by the 3GPP standards by adding an offset to the SSDL of small cells, to artificially strengthen the SSDL of small cells so that they are selected more often by the end-user devices. The BR provides a location-based solution to this problem, which can even be combined with the location information, where data have been received by end-user devices. With this combination even areas where only a few data packets can be received by the end-user devices through small cells can be identified and macro cells can be used instead of small cells. Additionally, the end-user device speed can be taken into account to prevent end-user devices from connecting to small cells at a high speed.

To prove the viability of the BR architecture, the fundamental parts have been implemented in the ns-3 network simulator. The simulation scenarios, the simulation setup and the results and its evaluation have been described and discussed in chapter 8. The simulations showed that the BR is able to improve all the values analysed and show better results than the reference simulation scenario. The simulation has proven the viability of the BR architecture and showed the benefit of the application of the BR.

Several papers covering different aspects of the proposed architecture, functionalities, and achieved results in this research have been presented at refereed conferences and published in a book and received positive comments from delegates and reviewers. A list of the published outputs can be found in the Appendix C.

9.2 Limitations of the Research

Despite having met the objectives of the research project, a number of decisions had to be made, which resulted in limitations on the work. These decisions were caused by practical, financial, and infrastructural access reasons. The limitations are summarised below.

1. The simulation implementation was restricted to the basic principle: implement as much functionality as necessary to prove that the defined architecture and its functionalities are viable and that the BR architecture brings benefit to fundamental use cases. This restriction was caused by the given timeline. Therefore, the simulation implementation considers only the most important parameters. The potential KPIs described for the BR architecture together with the introduced data gathering and distribution protocols are not all included in the BR implementation. Instead of protocol implementations for data gathering and the distribution of commands in the ns-3 network simulator, abstractions of the protocols have been used, to directly provide parameters to the BR or to distribute commands to the end-user device directly. The reason for this abstraction was to reduce the protocol implementation effort in ns-3, which was not in the scope of the simulation, because these standardised protocols have been proved through the standardisation organisations.

2. The BR is located in the cloud, which is in fact located at datacentres. In order to be able to integrate the BR into a cloud, a virtual server would have been necessary. Due to the financial and infrastructural access restrictions it was not possible to perform tests with a virtualised BR. The performance of cloud technology is a research item of its own and was not in the focus of this research.

Despite these limitations, the research project has led to valid contributions to knowledge on novel architectural and functional extensions of traffic steering mechanisms for mobile networks and provided sufficient proof-of-concept for the approaches proposed.

9.3 Directions of Further Research

This research project has advanced the field of individual traffic steering within heterogeneous mobile networks. However, a number of areas for future work can be identified, which are based on the results of this research and more generally within the area of traffic steering within heterogeneous mobile networks. These suggestions for future work are detailed below.

1. Additional effort could be put into completion of the ns-3 simulation to provide more parameters that can be used as KPIs to make the decision making process on handover and offload even smarter and more intelligent. The more parameters/KPIs are considered, the more precise the decision can be made.
2. Further ns-3 implementation work could be done in the field of the external modules. Additional external modules could be implemented and the impact of the coordination of these external modules could be analysed to evaluate and identify which combinations of external modules can bring benefits for which use cases.
3. The implementation of the several data gathering protocols and the protocol to distribute commands and information to the end-user device is very complex and requires adequate manpower. A partial implementation of the most important functions combined with abstractions made on the protocols, as it has already been done seems to be the practicable way. It is almost impossible that an outstanding researcher gets access to a test laboratory of a

real heterogeneous mobile network to integrate the BR directly into the laboratory environment and use existing interfaces that implement the protocols. The partial and abstracted protocol implementation of the main functionalities in the network simulator ns-3 is recommended and should be extended.

4. Additional ns-3 simulations could be performed with other mobility models, such as steady state random walk with variable end-user device speeds.
5. The capability of the Back Rider to support 5G mobile networks is provided with this thesis, but the evolvement towards 5G mobile networks has to be observed carefully. Certainly, there will arise more use cases than the carrier aggregation over Wi-Fi and LTE or the LTE/Wi-Fi link aggregation where the BR can be applied or adapted to, to bring benefit to new use cases and application areas.
6. With the emerging of the 5G technology the development of Software Defined Networking (SDN) as well as Network Function Virtualisation (NFV) is tightly coupled. It has to be analysed how the Black Rider fits into these new technologies. At first the BR can be deployed as a Virtual Network Function (VNF) in the cloud. The development on the end-user devices goes into the direction of providing a virtual router. This development opens up new possibilities to push commands and recommendations from the Black Rider towards the end-user device. Instead of using the S14 interface, the commands and recommendations could be pushed to the end-user devices with the SDN functions from the Black Rider directly to the virtual router on the end-user device. This implies that the Black Rider is implementing an SDN controller. These new possibilities should be addressed in future research.

References

1. 3GPP TR 22.985 V10.0.0, 2011. *Service requirements for the User Data Convergence (UDC)*. Release 10.
2. 3GPP TR 23.829 V1.3.0, 2010. *Local IP Access and Selected IP Traffic Offload*. Release 10.
3. 3GPP TR 23.834 V10.0.0, 2010. *Study on General Packet Radio Service (GPRS) Tunelling Protocol (GTP) based S2b*. Release 10.
4. 3GPP TR 23.852 V12.0.0, 2013. *Study on S2a Mobility based on GPRS Tunnelling Protocol (GTP) and Wireless Local Area Network (WLAN) access to the Enhanced Packet Core (EPC) network (SaMOG)*. Release 12.
5. 3GPP TR 23.859 V12.0.1, 2013. *Local IP access (LIPA) mobility and Selected IP Traffic Offload (SIPTO) at the local network*. Release 12.
6. 3GPP TR 23.861 V1.7.0, 2012. *Network based IP flow mobility*. Release 12.
7. 3GPP TR 23.865 V12.1.0, 2013. *Study on Wireless Local Area Network (WLAN) network selection for 3GPP terminals*. Release 12.
8. 3GPP TR 36.839 V11.1.0, 2012. *Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility enhancements in heterogeneous networks*. Release 11.
9. 3GPP TR 36.839, 2012. *Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility enhancements in heterogeneous networks*. Release 11: V11.1.0.
10. 3GPP TR 37.834 V12.0.0, 2013. *Study on Wireless Local Area Network (WLAN) - 3GPP radio interworking*. Release 12.
11. 3GPP TS 23.003 V12.1.0, 2013. *Numbering, addressing and identification*. Release 12.
12. 3GPP TS 23.203 V12.3.0, 2013. *Policy and charging control architecture*. Release 12.
13. 3GPP TS 23.228 V12.3.0, 2013. *IP Multimedia Subsystem (IMS)*. Release 12.

14. 3GPP TS 23.335 V10.0.0, 2011. *User Data Convergence (UDC); Technical realization and information flows*. Release 10.

15. 3GPP TS 23.401 V12.3.0, 2013. *General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access*. Release 12.

16. 3GPP TS 23.402 V12.3.0, 2013. *Architecture enhancements for non-3GPP accesses*. Release 12.

17. 3GPP TS 24.007 V12.0.0, 2013. *Mobile radio interface signalling layer 3*. Release 12.

18. 3GPP TS 24.302 V12.3.0, 2013. *Access to the 3GPP Evolved Packet Core (EPC) via non-3GPP access networks*. Release 12.

19. 3GPP TS 24.312 V12.3.0, 2013. *Access Network Discovery and Selection Function (ANDSF) Management Object (MO)*. Release 12.

20. 3GPP TS 29.212 V12.3.0, 2013. *Policy and Charging Control (PCC); Reference points*. Release 12.

21. 3GPP TS 29.213 V12.2.0, 2013. *Policy and Charging Control signalling flows and Quality of Service (QoS) parameter mapping*. Release 12.

22. 3GPP TS 29.214 V12.2.0, 2013. *Policy and Charging Control over Rx reference point*. Release 12.

23. 3GPP TS 29.274 V12.3.0, 2013. *Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C)*. Release 12.

24. 3GPP TS 29.281 V11.6.0, 2013. *General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U)*. Release 11.

25. 3GPP TS 36. Series (2011) *LTE (Evolved UTRA) and LTE-Advanced radio technology*, [Online], Available: http://www.3gpp.org/DynaReport/36-series.htm.

26. 3GPP TS 36.304 V11.6.0, 2013. *User Equipment (UE) procedures in idle mode*. Release 11.

27. 3GPP TS 36.331, 2014. *Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification*. Release 12.

28. 3GPP, TR 23.853 V12.0.0, 2012. *Operator Policies for IP Interface Selection (OPIIS)*. Release 12.

29. ABI Research (2012) *LTE Subscriber Totals Have Surpassed WiMAX in 2Q12*, [Online], Available: https://www.abiresearch.com/press/lte-subscriber-totals-have-surpassed-wimax-in-2q12 [27 February 2014].

30. ABI Research (2013) *9.7 million Carrier Wi-Fi Access Point Shipments in 2018 as Mobile Carriers Jump on the Bandwagon*, [Online], Available: https://www.abiresearch.com/press/97-million-carrier-wi-fi-access-point-shipments-in [27 February 2014].

31. ABI Research (2013) *Growing Demand for Mobility will Boost Global Wi-Fi Hotspots to Reach 6.3 Million in 2013*, [Online], Available: https://www.abiresearch.com/press/growing-demand-for-mobility-will-boost-global-wi-f [27 February 2014].

32. Aruba (2008) *Aruba Test*, [Online], Available: http://bradreese.com/aruba-test.pdf [June 2014].

33. Baldo, N., Miozzo, M., Requena-Esteso, M. and Nin-Guerrero, J. (2011) 'An open source product-oriented LTE network simulator based on ns-3', Proceedings of the 14th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems (MSWiM 11), New York, USA, 293-298.

34. Baldo, N., Requena-Esteso, M., Núñez-Martínez, J., Portolès-Comeras, M., Nin-Guerrero, J., Dini, P. and Mangues-Bafalluy, J. (2010) 'Validation of the IEEE 802.11 MAC Model in the ns3 Simulator using the EXTREME Testbed', Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques (SIMUTools), Torremolinos, Malaga (Spain).

35. Budiarto, H., Petrov, D., Puttonen, J. and Kurjenniemi, J. (2013) 'A3-Based Measurements and Handover Model for NS-3 LTE', The Third International Conference on Mobile Services, Resources, and Users, Lisbon, Portugal.

36. Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) (2014) *LTE Module*, [Online], Available: http://www.nsnam.org/docs/models/html/lte.html [August 2014].

37. Chang, B., Chen, J., Hsieh, C. and Liang, Y. (2009) 'Markov Decision Process-Based Adaptive Vertical Handoff with RSS Prediction in Heterogeneous Wireless Networks', Wireless Communications and Networking Conference (WCNC).

38. Cisco, 2014. *802.11ac: The Fifth Generation of Wi-Fi*. Technical White Paper.

39. Cisco, 2014. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018*.

40. ETSI TS 102 250-2, 2011. *Speech and multimedia Transmission Quality (STQ); QoS aspects for popular services in mobile networks*. ver. 2.2.1.

41. Fielding, R. (2000) *Architectural Styles and the Design of Network-based Software Architectures*, [Online], Available: http://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf [April 2014].

42. Frei, S., Fuhrmann, W., Rinkel, A. and Ghita, B.V. (2010) 'Signalling Effort Evaluation of Mobility Protocols within Evolved Packet Core Network', Proceedings of the Eighth International Network Conference (INC 2010), Heidelberg, Germany, 99-108.

43. Frei, S., Fuhrmann, W., Rinkel, A. and Ghita, B.V. (2011b) 'Improvements to Inter System Handover in the EPC Environment', Proceedings of the 4th IFIP International Conference on new Technologies, Mobility and Security (NTMS 2011), Paris, France, 1-5.

44. Frei, S., Fuhrmann, W., Rinkel, A. and Ghita, B.V. (2011c) 'Recent EPS Implementations Using ns-3', Workshop on Simulation and Prototyping Environments for Mobile/Wireless Research, 13. July 2011, Aachen, Aachen, Germany, http://itg.lkn.ei.tum.de/lib/exe/fetch.php?media=archiv:2011_07_13_aachen:15_itg524_aachen_frei.pdf.

45. Frei, S., Fuhrmann, W., Vergakis, D. and Rinkel, A. (2012b) 'Simulation Environment for the Evolved Packet System', 17. ITG Fachtagung Mobilkommunikation, Osnabrück, Germany, pp 83-88, ISBN: 987-3-8007-3438-2.

46. Gajic, B., Palenzuela, J., Riihijarvi, J. and Mahonen, P. (2012) 'Mobility-Aware Topology Manager for Publish-Subscribe Networks', Proceedings of the 5th International Conference on New Technologies, Mobility and Security (NTMS).

47. Ghinamo, G., Vadala, F., Corbi, C., Bettassa, P., Risso, F. and Sisto, R. (2012) 'Vehicle navigation service based on real-time traffic information: A RESTful NetAPI solution with long polling notification', Ubiquitous Positioning, Indoor Navigation, and Location Based Service (UPINLBS), 2012, 1-8.

48. GNU Octave (2014) *GNU Octave*, July, [Online], Available: http://www.gnu.org/software/octave/ [July 2014].

49. Gnuplot (2014) *Gnuplot*, July, [Online], Available: http://www.gnuplot.info/ [July 2014].

50. Gondara, M. and Kadam, S. (2011) 'Requirements of Vertical Handoff Mechanism in 4G Wireless Networks', *International Journal of Wireless & Mobile Networks (IJWMN), Vol. 3, No. 2*, April, pp. 18-27.

51. Huang, J., Qian, F., Gerber, A., Mao, M., Sen, S. and Spatscheck, O. (2012) 'A Close Examination of Performance and Power Characteristics of 4G LTE Networks', Proceedings of the 10th International Conference on Mobile Systems, Applications and Services MobiSys 2012, United Kingdom.

52. IEEE 1900.4.1, 2013. *IEEE Standard for Interfaces and Protocols Enabling Distributed Decision Making for Optimized Radio Resource Usage in Heterogeneous Wireless Networks*.

53. IEEE 1900.4, 2009. *IEEE Standard for Architectural Building Blocks Enabling Network-Device Distributed Decision Making for Optimized Radio Resource Usage in Heterogeneous Wireless Access Networks*.

54. IEEE 1900.4a, 2011. *Amendment 1: Architecture and Interfaces for Dynamic Spectrum Access Networks in White Space Frequency Bands*.

55. IEEE 802.11-2007, 2007. *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*.

56. IEEE 802.11-2012, 2012. *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*.

57. IEEE 802.11ac, 2013. *Amendment 4: Enhancements for Very High Throughput for Operation in Bands below 6 GHz*.

58. IEEE 802.11e, 2005. *Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements*.

59. IEEE 802.11k, 2008. *Amendment 1: Radio Resource Measurement of Wireless LANs*.

60. IEEE 802.11n, 2009. *Amendment 5: Enhancements for Higher Throughput*.

61. IEEE 802.11r, 2008. *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications; Amendment 2: Fast Basic Service Set (BSS) Transition*.

62. IEEE 802.11u, 2011. *Amendment 9: Interworking with External*.

63. IEEE 802.16, 2009. *Part 16: Air Interface for Broadband Wireless Access Systems*.

64. IEEE 802.16e, 2006. *Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands*.

65. IEEE 802.16m, 2011. *Amendment 3: Advanced Air Interface*.

66. IEEE 802.21, 2009. *IEEE Standard for Local and metropolitan area networks - Part 21: Media Independent Handover Services*.

67. IETF RFC 2460, 1998. *Internet Protocol, Version 6 (IPv6)*.

68. IETF RFC 2474, 1998. *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*.

69. IETF RFC 2679, 1999. *A One-way Delay Metric for IPPM*.

70. IETF RFC 2679, 1999. *A One-way Delay Metric for IPPM*.

71. IETF RFC 2680, 1999. *A One-way Packet Loss Metric for IPPM*.

72. IETF RFC 3393, 2002. *IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)*.

73. IETF RFC 3588, 2003. *Diameter Base Protocol*.

74. IETF RFC 3775, 2004. *Mobility Support in IPv6*.

75. IETF RFC 4006, 2005. *Diameter Credit-Control Application*.

76. IETF RFC 4960, 2007. *Stream Control Transmission Protocol*.

77. IETF RFC 5149, 2008. *Service Selection for Mobile IPv6*.

78. IETF RFC 5213, 2008. *Proxy Mobile IPv6*.

79. IETF RFC 5555, 2009. *Mobile IPv6 Support for Dual Stack Hosts and Routers*.

80. IETF RFC 5845, 2010. *Generic Routing Encapsulation (GRE) Key Option for Proxy Mobile IPv6*.

81. IETF RFC 768, 1980. *User Datagram Protocol*.

82. IETF, NETEXT Working Group, 2013. *Logical Interface Support for multi-mode IP Hosts*. draft-ietf-netext-logical-interface-support-08.

83. IETF, NETEXT Working Group, 2013. *Proxy Mobile IPv6 Extensions to Support Flow Mobility*. draft-ietf-netext-pmipv6-flowmob-08.

84. ISO/IEC 11172-3, 1993. *Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Part 3: Audio*.

85. ISO/IEC 13818-3, 1998. *Information technology -- Generic coding of moving pictures and associated audio information -- Part 3: Audio*.

86. ITU-R M.2134, 2008. *Requirements related to technical performance for IMT-Advanced radio interface(s)*.

87. ITU-T, 1988. *G7.11 Pulse code modulation (PCM) of voice frequencies*.

88. ITU-T, 2014. *H.264: Advanced video coding for generic audiovisual services*. (V9).

89. Jeong, B., Shin, S., Jang, I., Sung, N. and Yoon, H. (2011) 'A Smart Handover Decision Algorithm Using Location Prediction for Hierarchical Macro/Femto-Cell Networks', Vehicular Technology Conference (VTC Fall).

90. Jianfeng, L. and Jianglong, H. (2012) 'A RESTful information service method in Hybrid Sensor and Vehicular Networks', International Conference on Automatic Control and Artificial Intelligence (ACAI 2012), 283 - 286.

91. Kassar, M., Kervella, B. and Pujolle, G. (2008) 'An overview of vertical handover decision strategies in heterogeneous wireless networks', *Computer Communications, Elesevier*, June, pp. 2607-2620.

92. Louta, M. and Bellavist, P. (2013) 'Bringing Always Best Connectivity Vision a Step Closer: Challenges and Perspectives', *IEEE Communications Magazine*, February, pp. 158-166.

93. Mizouni, R., Serhani, M.A., Dssouli, R., Benharref, A. and Taleb, I. (2011) 'Performance Evaluation of Mobile Web Services', Ninth IEEE European Conference on Web Services (ECOWS), 184 - 191.

94. Munoz, P., Barco, R., Laselva, D. and Mogensen, P. (2013) 'Mobility-Based Strategies for Traffic Steering in Heterogeneous Networks', *IEEE Communications Magazine*, May, pp. 54-62.

95. Navidi, W. and Camp, T. (2004) 'Stationary Distributions for the Random Waypoint Mobility Model', *IEEE Transactions on Mobile Computing*, January-March, pp. 99-108.

96. Navidi, W., Camp, T. and Bauer, N. (2004) 'Improving the Accuracy of Random Waypoint Simulations Through Steady-State Initialization', Proceedings of the 15th International Conference on Modeling and Simulation (MS '04), 319-326.

97. ns-3 community (2014a) *WiFi*, [Online], Available: http://www.nsnam.org/docs/models/html/wifi.html# [August 2014].

98. ns-3 community (2014b) *ns-3 Model Library*, [Online], Available: http://www.nsnam.org/docs/models/html/index.html [August 2014].

99. OMA (2011a) *DM DiagMon Architecture, Approved Version 1.0, OMA-AD-DM-DiagMon-V1_0-20111220-A*, 20 December, [Online], Available: http://technical.openmobilealliance.org/Technical/release_program/docs/DiagMon/V1_0-20120313-A/OMA-AD-DiagMon-V1_0-20111220-A.pdf [April 2014].

100. OMA (2011b) *OMA DiagMon Management Object V1.1*, [Online], Available: http://technical.openmobilealliance.org/Technical/release_program/diagmon_V1_1.aspx [May 2014].

References

101. OMA (2013a) *Enabler Release Definition for OMA Device Management Version 2.0*, [Online], Available: http://technical.openmobilealliance.org/Technical/technical-information/release-program/current-releases/oma-device-management-v2-0 [April 2014].

102. OMA (2013b) *OMA DiagMon Management Object V1.2*, [Online], Available: http://technical.openmobilealliance.org/Technical/release_program/diagmon_v1_2.aspx [May 2014].

103. Osseiran, A., Boccardi, F., Braun, V., Kusume, K., Marsch, P., Maternia, M., Queseth, O., Schellmann, M., Schotten, H., Taoka, H., Tullberg, H., Uusitalo, M.A., Timus, B. and Fallgren, M. (2014) 'Scenarios for 5G Mobile and Wireless Communications: The Vision of the METIS Project', *Communications Magazine, IEEE*, Issue: 5 Volume: 52, pp. 26-35.

104. Pedersen, K.I., Michaelsen, P.H., Rosa, C. and Barbera, S. (2013) 'Mobility Enhancements for LTE-Advanced Multilayer Networks with Inter-Site Carrier Aggregation', *IEEE Communications Magazine*, May, pp. 64-71.

105. Pei, G. and Henderson, T.R. (2010) 'Validation of OFDM Error Rate Model in ns-3', *Boeing Research Technology*, pp. 1-15.

106. Poese, I., Uhlig, S., Kaafar, M.A., Donnet, B. and Gueye, B. (2011) 'IP geolocation databases: unreliable?', *ACM SIGCOMM Computer Communication Review, Volume 41 Issue 2*, April, pp. 53-56.

107. Qualcomm, 2011. *A Comparison of LTE Advanced HetNets and Wi-Fi*.

108. RootMetrics (2014) *RootMetrics.com*, [Online], Available: http://www.rootmetrics.com [June 2014].

109. Sesia, S., Toufik, I. and Baker, M. (2011) *LTE - The UMTS Long Term Evolution: From Theory to Practice*, 2nd edition, ISBN: 978-0-470-66025-6: Wiley.

110. Souders, S. (2014) *HTTP Archive*, [Online], Available: http://www.httparchive.org [June 2014].

111. Szigeti, T. and Hattingh, C. (2004) *End-to-End QoS Network Design: Quality of Service in LANs, WANs, and VPNs*, Cisco Press.

112. Tervonen, J. and Mustajärv, J. (2010) *Realization of Policy-Based Resource Management Concept, Future Internet program of TIVIT*, 16 February, [Online], Available: http://www.futureinternet.fi/publications/DA2.2.20_v1.0.pdf [30 April 2014].

113. Wi-Fi Alliance Hotspot 2.0, 2012. *Hotspot 2.0 Technical Specification*. Version 1.0.0, Release 1.

114. XIRRUS, 2013. *802.11ac Demystified*. White paper.

115. Zink, M., Suh, K., Gu, Y. and J. Kurose (2008) 'Watch Global, Cache Local: YouTube Network Traffic at a Campus Network - Measurements and Implications', Proceedings of SPIE/ACM Conference on Multimedia Computing and Networking (MMCN) 2008, Santa Clara, USA.

Appendices

A Energy Consumption Use Case Calculations

A.1 High Energy Consumption With and Without Black Rider

The energy consumption is for both scenarios, with and without the Black Rider, the same, because if the High Energy Consumption (HEC) energy model is applied the energy consumption is irrelevant. The focus is then on high performance.

Used RAT: LTE									
Application	**Duration Total [s]**	**Duration UL [s]**	**Duration DL [s]**	**Total [Mbps]**	**total traffic UL [Mbps]**	**total traffic DL [Mbps]**	**Pu [W]**	**Pd [W]**	**Total [Ws]**
Skype	720	360	360	46.08	23.04	23.04	1.316097	1.2913661	938.6866944
YouTube	450	0.01038961	449.9896104	346.5	0.008	346.492	1.2915471	1.3280569	597.6252257
Audio streaming	1200	0.0625	1199.9375	153.6	0.008	153.592	1.2915471	1.2946922	1553.630395
Idle	1230								730.989
Total Energy Consumption LTE							**HEC with and without BR [Ws]**		**3820.931316**

A.2 Medium Energy Consumption With Black Rider

RAT: LTE									
Application	**Total Duration [s]**	**Duration UL [s]**	**Duration DL [s]**	**Number**	**tu [Mbps]**	**td [Mbps]**	**Pu [W]**	**Pd [W]**	**Total [Ws]**
Web browsing	10.22556596	0.0128	10.21276596	14	0.008	12	1.29154712	1.91168	19.5400722
	Duration Total [s]	**Duration UL [s]**	**Duration DL [s]**	**Total [Mbps]**	**total traffic UL [Mbps]**	**total traffic DL [Mbps]**	**Pu [W]**	**Pd [W]**	**Total [Ws]**
Skype	600	300	300	38.4	19.2	19.2	1.31609696	1.29136608	782.238912
Used LTE time	610.225566								
LTE total	2640								

LTE Idle	2029.774434									1206.29495
Total Energy Consumption LTE										2008.07393
RAT: Wi-Fi										
Application	**Total Duration [s]**	**Duration UL [s]**	**Duration DL [s]**	**Number**	**tu [Mbps]**		**td [Mbps]**	**Pu [W]**	**Pd [W]**	**Total [Ws]**
Web browsing	55.95940803	0.145454545	55.81395349	20	0.008		4.3	0.13512536	0.722003	40.3174965
Send Email	**Total Duration [s]**	**Duration UL [s]**	**Duration DL [s]**	**Number**	**tu [Mbps]**		**td [Mbps]**	**Pu [W]**	**Pd [W]**	**Total [Ws]**
Email plain text	2.912811839	2.909090909	0.00372093	2	0.008		1.6	0.13512536	0.352076	0.39440201
Email with attachment	87.27830867	87.27272727	0.005581395	3	0.008		4.3	0.13512536	0.722003	11.7967885
Receive Email										
Email plain text	1.896828753	0.036363636	1.860465116	5	0.008		1.6	0.13512536	0.352076	0.65993877
Email with attachment	22.34739958	0.021818182	22.3255814	3	0.008		4.3	0.13512536	0.722003	0.00447836
Used Wi-Fi time	170.3947569									
Wi-Fi Total time	960									
Wi-Fi Idle	789.6052431									60.9575248
Total Energy Consumption Wi-Fi										114.130629
Total Energy Consumption LTE + Wi-Fi								**MEC with BR [Ws]**		**2122.20456**

A.3 Medium Energy Consumption Without Black Rider

RAT: LTE									
Application	**Total Duration [s]**	**Duration UL [s]**	**Duration DL [s]**	**Number**	**tu [Mbps]**	**td [Mbps]**	**Pu [W]**	**Pd [W]**	**Total [Ws]**
Web browsing	24.83351733	0.031085714	24.80243161	34	0.008	12	1.29154712	1.91168	47.4544611
Skype	600	300	300		19.2	19.2	1.31609696	1.29136608	782.238912
Send Email	**Total Duration [s]**	**Duration UL [s]**	**Duration DL [s]**	**Number**	**tu [Mbps]**	**td [Mbps]**	**Pu [W]**	**Pd [W]**	**Total [Ws]**
Email plain text	0.36668693	0.365714286	0.000972644	2	0.008	1.6	1.29154712	1.371192	0.47367091
Email with attachment	10.97288754	10.97142857	0.001458967	3	0.008	16.45	1.29154712	2.1429465	14.1732435
Receive Email									
Email plain text	0.490893617	0.004571429	0.486322188	5	0.008	1.6	1.29154712	1.371192	0.67274531
Email with attachment	5.838609119	0.002742857	5.835866261	3	0.008	16.45	1.29154712	2.1429465	12.5094917

Used LTE time	642.5025945			
LTE Total time	3600			
LTE Idle	2957.497405			1757.64071
Total Energy Consumption LTE			**MEC without BR [Ws]**	**2615.16323**

A.4 Low Energy Consumption With Black Rider

The used RATs for this scenario are 3G and Wi-Fi.

RAT: 3G

Application	**Total Duration [s]**	**Duration UL [s]**	**Duration DL [s]**	**Number**	**tu [Mbps]**	**td [Mbps]**	**Pu [W]**	**Pd [W]**	**Total [Ws]**
Web browsing	1.091636364	0.00072727	1.09090909	2	0.008	12	0.82483184	2.28332	2.49149442
Used 3G time	1.091636364								
3G total	2640								
3G Idle	2638.908364								987.47951
Total Energy Consumption 3G									989.971004

RAT: Wi-Fi

Application	**Total Duration [s]**	**Duration UL [s]**	**Duration DL [s]**	**Number**	**tu [Mbps]**	**td [Mbps]**	**Pu [W]**	**Pd [W]**	**Total [Ws]**
Web browsing	27.97970402	0.07272727	27.9069767	10	0.008	4.3	0.13512536	0.722003	20.1587482
Send Email	**Total Duration [s]**	**Duration UL [s]**	**Duration DL [s]**	**Number**	**tu [Mbps]**	**td [Mbps]**	**Pu [W]**	**Pd [W]**	**Total [Ws]**
Email plain text	4.369217759	4.36363636	0.0055814	3	0.008	1.6	0.13512536	0.352076	0.59160301
Email with attachment	29.09276956	29.0909091	0.00186047	1	0.008	4.3	0.13512536	0.722003	0.0010923
Receive Email					0.008				
Email plain text	1.896828753	0.03636364	1.86046512	5	0.008	1.6	0.13512536	0.352076	0.65993877
Email with attachment	22.34739958	0.02181818	22.3255814	3	0.008	4.3	0.13512536	0.722003	16.1220849
Used Wi-Fi time	61.44169133								
Wi-Fi Total time	960								
Wi-Fi Idle	898.5583087								69.3687014
Total Energy Consumption Wi-Fi									106.902169
Total Energy Consumption 3G + Wi-Fi								**LEC with BR [Ws]**	**1096.87317**

A.5 Low Energy Consumption Without Black Rider

RAT: 3G									
Application	**Total Duration [s]**	**Duration UL [s]**	**Duration DL [s]**	**Number**	**tu [Mbps]**	**td [Mbps]**	**Pu [W]**	**Pd [W]**	**Total [Ws]**
Web browsing	6.549818182	0.00436364	6.54545455	12	0.008	12	0.82483184	2.28332	0.00415249
Send Email	**Total Duration [s]**	**Duration UL [s]**	**Duration DL [s]**	**Number**	**tu [Mbps]**	**td [Mbps]**	**Pu [W]**	**Pd [W]**	**Total [Ws]**
Email plain text	4.369217759	4.36363636	0.0055814	3	0.008	1.6	0.82483184	1.013272	3.60492168
Email with attachment	29.09276956	29.0909091	0.00186047	1	0.008	4.3	0.82483184	1.342996	23.9976067
Receive Email									
Email plain text	1.896828753	0.03636364	1.86046512	5	0.008	1.6	0.82483184	1.013272	1.91515109
Email with attachment	22.34739958	0.02181818	22.3255814	3	0.008	4.3	0.82483184	1.342996	30.0011628
Used 3G time	64.25603383								
3G Total time	3600								
3G Idle	3535.743966								1323.07539
Total Energy Consumption 3G								**LEC without BR [Ws]**	**1382.59423**

B Publications

B.1 Conference Papers

1. S. Frei, W. Fuhrmann, B.V. Ghita (2013b) 'Generic Real-Time Traffic Distribution Framework: Black Rider', 22nd International Conference on Computer Communications and Networks (ICCCN 2013), 30. July – 2. August 2013, Nassau, Bahamas; ISBN: 978-1-4673-5774-6, pp 1-8, DOI: 10.1109/ICCCN.2013.6614

2. S. Frei, W. Fuhrmann, B.V. Ghita (2013a) 'Framework for Generic Context- and Policy-Based Traffic Distribution in Heterogeneous Wireless Networks: Black Rider', 27th International Conference on Advanced Information Networking and Applications Workshops (WAINA), 25-28 March 2013, Barcelona, Spain; ISBN: 978-1-4673-6239-9, pp 534 – 541, DOI: 10.1109/WAINA.2013.99

3. S. Frei, W. Fuhrmann, A. Rinkel, B.V. Ghita (2012) 'Prospects for WLAN in the Evolved Packet Core Environment', 5th IFIP International Conference on new Technologies, Mobility and Security (NTMS 2012), 7-10 May 2012, Istanbul, Turkey; ISBN: 978-1-4673-0228-9, pp 1-5, DOI: 10.1109/NTMS.2012.6208674

4. S. Frei, W. Fuhrmann, D. Vergakis, A. Rinkel (2012) 'Simulation Environment for the Evolved Packet System', 17. ITG Fachtagung Mobilkommunikation, 9-10 May 2012, Osnabrück, Germany; ISBN: 987-3-8007-3438-2, pp 83-88

5. S. Frei, W. Fuhrmann, A. Rinkel, B.V. Ghita (2011c) 'EPS QoS Enforcement on Layer 3 with DiffServ', Proceedings of the Fourth International Conference on Internet Technologies and Applications (ITA 11), 6-9 Sept. 2011, Glyndwr University, Wrexham, UK; ISBN: 978-0-946881-68-0, pp 199-204

6. S. Frei, W. Fuhrmann, A. Rinkel, B.V. Ghita (2011b) 'Recent EPS Implementations Using ns-3', Workshop on Simulation and Prototyping Environments for Mobile/Wireless Research; 36th Meeting of VDE/ITG Working Group 5.2.4 Mobilität in IP-basierten Netzen, 13. July 2011, Aachen, Germany,
   http://itg.lkn.ei.tum.de/lib/exe/fetch.php?media=archiv:2011_07_13_aachen:15_itg524_aachen_frei.pdf

7. S. Frei, W. Fuhrmann, A. Rinkel, B. V. Ghita (2011a) 'Improvements to Inter System Handover in the EPC Environment', The 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS), IEEE Conference, 7–10 Feb. 2011, Paris, France; ISBN: 978-1-4244-8705-9, pp 1-5, DOI: 10.1109/NTMS.2011.5720589

8. S. Frei, W. Fuhrmann, A. Rinkel, B. V. Ghita (2010b) 'QoS enforcement within the EPS considering the macro mobility protocols', 2nd Next Generation Network Workshop, 2. Nov. 2010, Leipzig, Germany, Online: http://ngnlab.eu

9. S. Frei, W. Fuhrmann, A. Rinkel, B. V. Ghita (2010a) 'Signalling Effort of Mobility Protocols within Evolved Packet Core Network', Proceedings of the Eighth International Network Conference (INC 2010), 6–8 July 2010; Heidelberg, Germany, ISBN: 978-1-84102-259-8, pp 99-108

10. S. Frei, W. Fuhrmann, A. Rinkel, B. V. Ghita (2009) 'Reference Architecture for End-to-End QoS in Heterogeneous Wireless Network Environments', Proceedings of the Third International Conference on Internet Technologies and Applications (ITA 09), 8–11 September 2009, Glyndwr University, Wrexham, UK, ISBN: 978-0-946881-65-9, pp 630-639

B.2 Book

11. S. Frei (2008) 'Entwicklung einer WLAN Experimentierplattform: Zur Untersuchung, Planung und Bewertung von WLAN Zugängen', VDM Verlag Dr. Müller, Oktober 2008; ISBN: 978-3-639-08942-4

B.3 Poster

12. F. Mauchle, S. Frei, A. Rinkel (2010) 'Simulating Mobile IPv6 with ns-3', 3rd International ICST Conference on Simulation Tools and Techniques (SimuTools 2010),15–19 March 2010, Torremolinos, Malaga, Spain, ISBN 978-963-9799-87-5, DOI: 10.4108/ICST.SIMUTOOLS2010.8682

B.4 Internal Publications

13. S. Frei, W. Fuhrmann, A. Rinkel, B. V. Ghita (2009) 'Bottom up survey of the WiMAX technology', Proceedings of the Fifth Collaborative Research

Symposium on Security, E-learning, Internet and Networking (SEIN 2009), 25–29 Nov. 2009, Darmstadt, Germany, ISBN: 978-1-84102-236-9, pp 160-172

14. S. Frei, W. Fuhrmann, A. Rinkel, B. V. Ghita (2008) 'End-to-End QoS and Mobility in Wireless Access Networks Interworking with the 3GPP EPC', Proceedings of the Fourth Collaborative Research Symposium on Security, E-learning, Internet and Networking (SEIN 2008), 5–9 Nov. 2008, Glyndwr University, Wrexham, UK, ISBN: 978-1-84102-196-6, pp 194-207

B.5 Copies of Publications

The copies of publications have been removed due to Copyright restrictions.